REPORT DOCUMENTATION PAGE	Form Approved OMB No 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	March 1993	Master's Thesis

4. TITLE AND SUBTITLE

Feature Based Neural Network Acoustic Transient Signal Classification

5. FUNDING NUMBERS

6. AUTHOR(S)

Terry S. Wichert

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Postgraduate School

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

THE VIEWS EXPRESSED IN THIS THESIS ARE THOSE OF THE AUTHOR AND DO NOT REFLECT THE OFFICIAL POLICY OR POSITION OF THE DEPARTMENT OF DEFENSE OR THE U.S. GOVERNMENT.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Utilization of neural network techniques to recognize and classify acoustic signals has long been pursued and shows great promise as a robust application of neural network technology. Traditional techniques have proven effective but in some cases are quite computationally intensive, as the sampling rates necessary to capture the transient result in large input vectors and thus large neural networks. This thesis presents an alternative transient classification scheme which considerably reduces neural network size and thus computation time. Parameterization of the acoustic transient to a set of distinct characteristics (e.g. frequency, power spectral density) which capture the structure of the input signal is the key to this new approach. Testing methods and results are presented on networks for which computation time is a fraction of that necessary with traditional methods, yet classification reliability is maintained. Neural network acoustic classification systems utilizing the above techniques are compared to classic time domain classification networks. Last, a case study is presented which looks at these techniques applied to the active acoustic intercept problem.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Neural Networks, Antisubmarine Warfare, Acoustic Transient Classification, Active Acoustic Intercept	111
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500 Standard Form 298 (Rev. 2-89) Prescribed by ANSI Std. Z39-18

Approved for public release; distribution is unlimited.

Feature Based Neural Network Acoustic Transient Signal Classification

by

Terry S. Wichert
Lieutenant Commander, United States Navy
B.S., Oklahoma State University

Submitted in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE IN APPLIED PHYSICS

from the

NAVAL POSTGRADUATE SCHOOL
March 1993

Department of Physics

ABSTRACT

Utilization of neural network techniques to recognize and classify acoustic signals has long been pursued and shows great promise as a robust application of neural network technology. Traditional techniques have proven effective but in some cases are quite computationally intensive, as the sampling rates necessary to capture the transient result in large input vectors and thus large neural networks. This thesis presents an alternative transient classification scheme which considerably reduces neural network size and thus computation time. Parameterization of the acoustic transient to a set of distinct characteristics (e.g. frequency, power spectral density) which capture the structure of the input signal is the key to this new approach. Testing methods and results are presented on networks for which computation time is a fraction of that necessary with traditional methods, yet classification reliability is maintained. Neural network acoustic classification systems utilizing the above techniques are compared to classic time domain classification networks. Last, a case study is presented which looks at these techniques applied to the acoustic intercept problem.

C. 1

TABLE OF CONTENTS

I. INTRODUCTION

A. TRADITIONAL PROCESSING

The purpose of this thesis is to present a new method for classifying extremely short duration unintentional acoustic transients, utilizing neural network computing methods. This thesis presents an acoustic transient classification scheme which serves to take advantage of the inherent feature extraction capability of neural networks.

An acoustic transient is a transient wave which results from the sudden release of energy associated with any of a large number of events in the ocean environment. Examples include the snapping of the tail of a shrimp against its body as it seeks to propel itself, the rattle of two links of chain tethering a navigation buoy, and the stress incurred or released as the metal hull of a submarine is compressed or expanded during changes in depth. These types of transients are detectable with underwater pressure sensitive hydrophones but are often very difficult if not impossible to classify, owing to extremely short signal duration.

Traditional acoustic transient signal analysis has relied on classic techniques of Fourier analysis. See Figure 1. These generally include sensing the analog signal, sampling the signal at some rate (typically just above the Nyquist rate), feeding the now discrete signal to a Fast Fourier Transform

(henceforth referred to as FFT) machine, analyzing the signal for frequency content, and finally comparing the signal against the characteristics of signals known to contain similar frequency content.

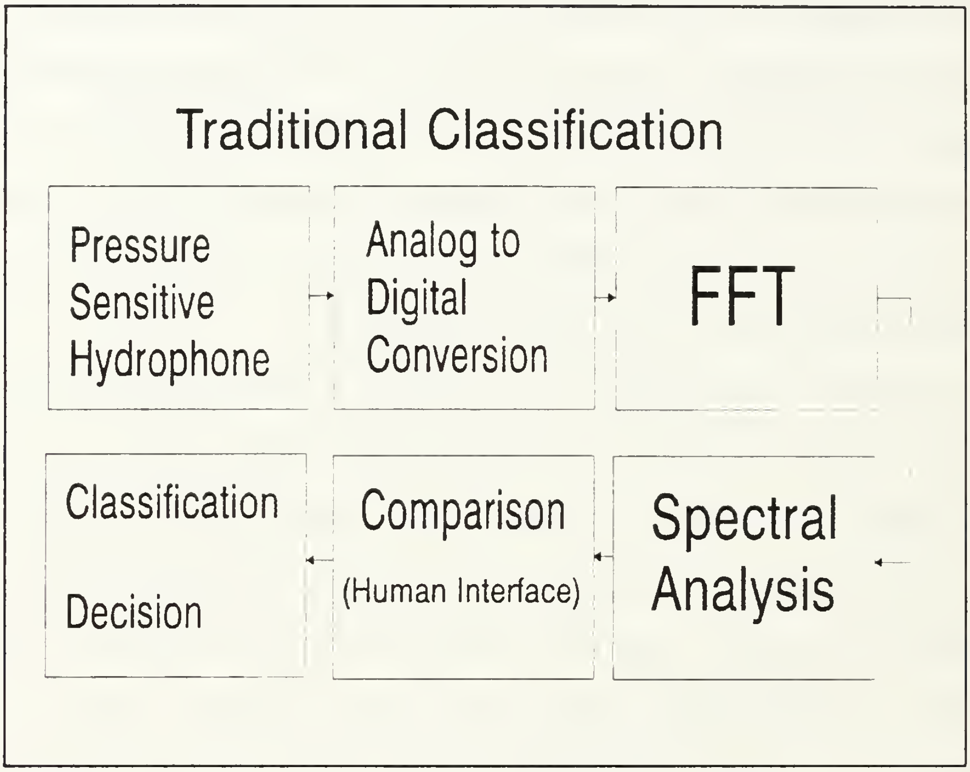

Figure 1: Traditional Signal Classification

These techniques have proven to be feasible, although somewhat computationally intensive, for continuous analog and moderate duration transient acoustic signals.

B. NEURAL PROCESSING

In recent years neural networks have offered an alternative approach to pattern recognition and signal processing based on automated learning procedures. Neural networks are attractive as a means of classifying acoustic

transients because they are capable of discovering features and patterns of interrelated features which serve to define the corresponding class of a signal. Additionally this method of pattern classification is desirable because a neural network has an ability to learn this structure and thus is capable of generalizing to novel or new but similar patterns. This being said, most neural network researchers in this area have attempted to utilize time series data or its Fourier transformed frequency counterpart directly as input to the network classifier. This approach is certainly advantageous when viewed in light of the arguments previously suggested and when compared to the computation time and reliability of the systems utilizing methods displayed in Figure 1. However this method is not without difficulties of its own. Foremost among problems associated with this type of approach is the need to "find" and extract the transient within a much larger data field and then to properly center the data prior to presentation to the network. Others have studied this problem and a good discussion of workable extraction methods is contained in a master's thesis by Shipley [Ref. 1]. Additionally given that the extraction has been made successfully the resulting input data vector can itself be quite large, which of course leads to a larger neural network and thus longer computation time. As an example suppose that a 10 msec duration transient containing frequencies in the range 3-10 kHz is to be detected. By the Nyquist sampling

theorem:

$$f_s = 2 \cdot f_{max} \tag{1}$$

Where

f_s= The sampling frequency

f_{max} = The maximum frequency contained within the signal

The sampling frequency for this case is 20 kHz. Sampled over 10 msec this results in 200 data points, necessitating a neural network input layer of 200 units and perhaps a total network size of 300 units. Although not computationally unreasonable by today's computing standards this thesis proposes to show that this same signal can be reliably classified with a neural network utilizing less than 40 units. Additionally the methods presented here do not suffer from many of the limitations outlined above. Namely there is no need to center data and remarkably network size is independent of signal duration. Figure 2 represents a conceptual block overview of the classification process described herein. This method stands in sharp contrast to that realized by classical methods such as those outlined in Figure 1. Note for example that although signal pre-processing is required, the human interface is gone, having occurred prior to signal pre-processing, in a less demanding environment.

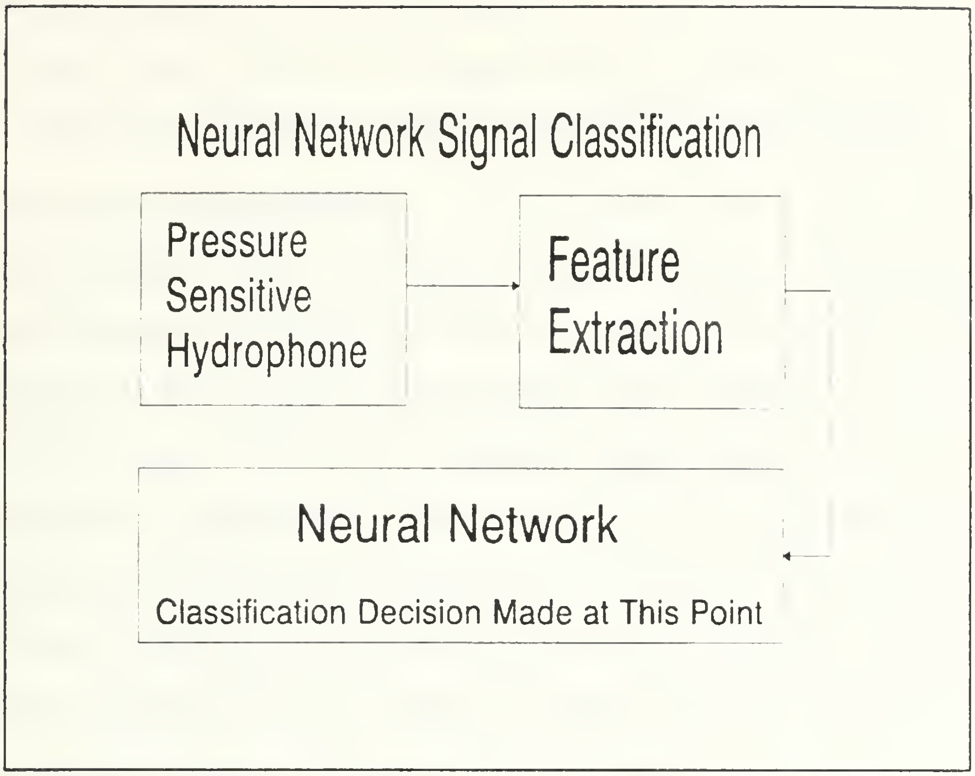

Figure 2: Neural Network Signal Classification

C. OBJECTIVES

This thesis produces a neural network transient acoustic signal classifier using commercially available software and hardware. This thesis utilizes data which has undergone signal pre-processing to parameterize the data into 31 individual features as input to the feature based neural classifier.

Further, this thesis compares the performance of this feature based classifier with time and frequency domain neural classifiers. Based on this comparison a feature based network which is considerably reduced in size is built, tested and analyzed.

Finally a case study is presented which demonstrates one

possible application of the neural computing analysis which is done in the balance of the thesis. In this case study the neural computing concepts and ideas presented herein are applied to the active acoustic intercept problem.

Elementary discussions of acoustic and neuralcomputing fundamentals as they relate to pattern recognition immediately follow this introduction. These should serve the uninitiated reader with enough neural network knowledge to comfortably read the remainder of the thesis. The remainder of the thesis is devoted to describing how the software tools were used to analyze the signals, how the data were analyzed using the neural network to prune down the size of the original feature based network, and side by side analysis of the new and traditional neural network transient detection methods emphasizing the results of how the smaller more efficient network performed.

II. ACOUSTIC AND NEURAL NETWORK FUNDAMENTALS

A. ACOUSTIC FUNDAMENTALS

This thesis deals primarily with signal processing of passive acoustic transient data. Although standard signal processing techniques exist for acoustic data, surprisingly little has been written on passive acoustic transient data. Thus some of the analysis overview presented here is borrowed from active sonar signal processing which by its very nature deals with the question of transient processing, namely the acoustic transient associated with the return of an active sonar emission from an acoustically reflective object.

When considering the processing of acoustic information in the ocean it is necessary to first consider the nature of sound in the ocean. The data analyzed in this thesis is transient noise produced from a moving source which is a fixed distance from a receiver which, in turn, listens through a background of noise. It is then relevant to look at the many difficulties associated with the detection of this signal.

The nature of the general passive acoustic problem is well documented [Ref. 2]. A classical argument is one in which a source and source level are defined. The many ways in which energy from the source is lost as the sound propagates through the ocean is then characterized. Finally the difficulties associated with detection of a signal in the presence of

background noise is quantified. Urick provides an excellent overview for the interested reader [Ref. 2].

Presented here is a specific discussion relevant to gathering and processing acoustic information in the ocean environment and a brief development of the nature of transients which allows direct substitution in the normal intensity based form of the passive sonar equations.

The data utilized in this thesis were gathered by a passive acoustic pressure based receiver listening in the noise laden ocean environment. The hydrophone, in its simplest form, is an electroacoustic transducer which measures the ambient pressure field directly through surface displacement and converts the field fluctuations to a voltage series in time through the piezoelectric effect. The user is provided then with a voltage series which represents the pressure field as a function of time at the receiver. Of course the hydrophone is calibrated before being placed in the water and thus the voltage series can readily be returned to a pressure field through:

$$V_x = M_{0x} P_T \qquad (2)$$

Where

V_x = Voltage recorded by the hydrophone

M_{0x}= The sensitivity of the hydrophone

P_T = The pressure field

This conversion is convenient for a number of reasons. First the pressure field can be processed to produce useful parametric measurements such as signal power, signal mass density, signal amplitude, etc. Most importantly, the signal can now be related to a Sound Pressure Level (SPL):

$$SPL=20\log\frac{P_e}{P_{ref}} \qquad (3)$$

Where

P_e=Effective Pressure = $P_T/(2)^{1/2}$

Last the voltage or pressure time series can be transformed to the frequency domain through standard FFT techniques and a whole new series of parametric information can be extracted, such as power spectral density, spectral moments, etc.

Now a short development of the acoustic nature of transients is presented as well as how these transients are transformed to relate them to the intensity based form of the passive sonar equations.

Typically the sonar equations are formulated in terms of intensity in the radiated sound field. A more general approach specific to the characterization of a transient is to write the equations in terms of energy flux density, defined as the acoustic energy per unit area of the transient wavefront, which is the time integral of the instantaneous intensity.

$$E=\int_0^\infty I dt=\frac{1}{\sigma c}\int_0^\infty p^2 dt \tag{4}$$

Where:

I = Intensity

c = Sound Speed

p = Acoustic pressure

σ = Density

In this case then the Intensity of the transient can be thought of as the mean square pressure of the wave divided by the specific acoustic impedance and averaged over an integral of time T:

$$I=\frac{1}{T}\int_0^T \frac{p^2(t)}{\sigma c} dt \tag{5}$$

The quantity T is often hard to define for short duration signals. However it can be shown that the intensity form of the sonar equations can be used, provided that the source level is defined as:

$$SL = 10\cdot\log(E)-10\cdot\log(\tau_e) \tag{6}$$

Where

SL= Source Level of the transformed transient

τ_e = the duration of the transient

This is convenient because it allows processing of short duration transients utilizing traditional methods of sonar signal processing. This type of processing will prove convenient for time series analysis. [Ref. 2]

As stated in the introduction this thesis is about <u>recognition</u> of acoustic information. Accordingly it is necessary to provide the reader with some basic fundamentals in what neural networks are and do. It is hoped that this overview will provide the uninitiated reader with sufficient knowledge to extract that which he finds relevant to his own particular interests and endeavors.

B. NEURAL NETWORK FUNDAMENTALS

This section serves to provide the reader with an introduction to neural network computing fundamentals which stands alone and will facilitate the discussions in the follow on sections.

In a strict formal sense a neural network is:

> "A parallel, distributed information processing structure consisting of processing elements (which can possess a local memory and can carry out localized information processing operations) interconnected via unidirectional signal channels called connections. Each processing element has a single output connection that branches ("fans out") into as many collateral connections as desired; each carries the same signal- the processing element output signal. The processing element output signal can be of any mathematical type desired. The information processing that goes on within each processing element can be defined arbitrarily with the restriction that it must be completely local; that is it must depend only on the current values of the input signals arriving at the processing element via impinging connections and on values stored in the processing element's local memory."[Ref. 3]

In a more practical sense a neural network consists of a computer architecture which incorporates all of the following:

1) A connection geometry for individual processing elements (henceforth referred to as neurons)

2) A transfer function which tells the network how to map or pass data from one neuron to others.

3) A learning rule which allows the network to improve its ability (learn by reducing error) to properly map the input to the output after repeated presentations of both.

4) An algorithm for minimizing output error.

1. CONNECTION GEOMETRIES

Connection geometries are simply the manner in which individual neurons are connected to facilitate the transfer of data. Figure 3 provides an example of one such geometry. The commonest type of artificial neural network consists of three layers of neurons. A layer of input neurons is connected to a layer of "hidden" neurons which is connected to a layer of output neurons. Although there is more than one way to connect this architecture, the networks considered in this thesis are all fully interconnected, i.e. each neuron in each layer is fully connected to each neuron in each layer immediately above and below it. Thus Figure 3 consists of one input layer with 6 neurons, one hidden layer with 3 neurons, and one output layer with 2 output neurons. All the neurons are fully interconnected as shown in the figure and discussed above. Also shown in Figure 3 is a bias unit. This bias unit acts

much like an electrical ground, maintaining a constant base level of activity when the activity of the neuron falls below a selectable threshold value.

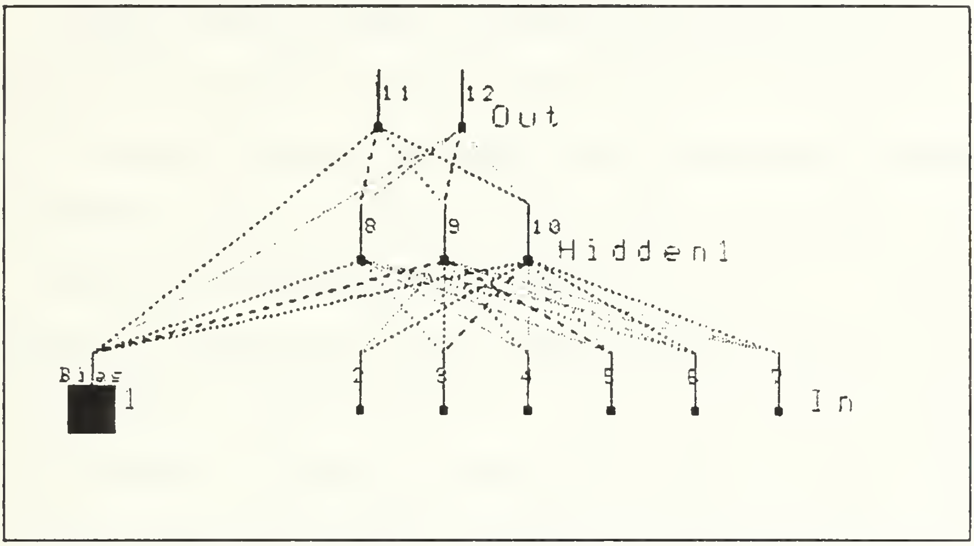

Figure 3: Typical Fully Connected Neural Network

2. TRANSFER FUNCTIONS

One important feature of neurocomputing with neural networks is the manner in which data is passed and manipulated between neurons of one layer and neurons of another layer and within the neuron itself. This process of manipulating data within the neuron is accomplished mathematically by use of a transfer function. This function uses local memory and input to the neuron to produce the activation level for the neuron. Essentially the transfer function receives inputs as values stored in local memory corresponding to the current state of the neuron and it also receives input via the connections to the neuron. The transfer function then performs a mathematical operation on the inputs and produces two quantities, namely

the output activation level of the neuron, i.e. that signal which is passed on to other neurons via connections at the next update, and an activation level which is stored in local memory and corresponds to the new state of the neuron.

Transfer functions can really be any of a variety of mathematical functions which provide proper operation of the network. Experience and experimentation has limited these practically in most cases to the sigmoid function, the hyperbolic tangent function and other trigonometric functions, and straight linear mapping. In practice the most widely used transfer function is the sigmoid function because of an ability to map the real numbers $(-\infty,\infty)$ to the set $(0,1)$. The work presented in this thesis was done with the sigmoid function as a mapping transfer function. The sigmoid function is defined as:

$$f(x) = \frac{1}{1 + e^{-a \cdot x}} \tag{7}$$

This function has the properties that it is a bounded differentiable real function. It is bounded and monotonic increasing for all real inputs and has a positive derivative everywhere. Further, it is essentially linear for input values which are near the central point of the function (input values near zero). These properties make it convenient for use in generalized delta rule learning which will be discussed in the next section. Figure 4 illustrates graphically these features

and demonstrates the concept of mapping a large range of inputs (-100,100) to a small range of outputs (0,1), one feature which makes it desirable as a transfer function.

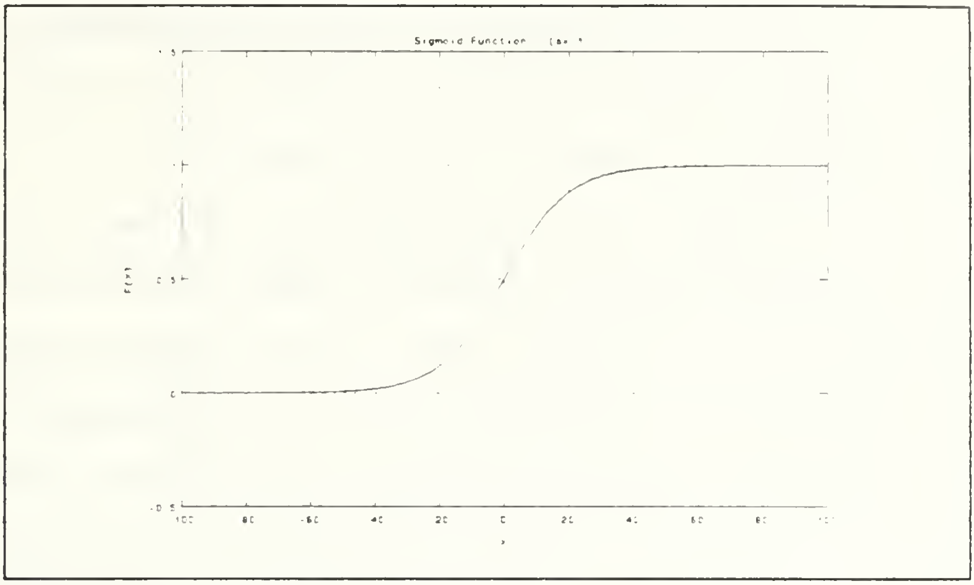

Figure 4: Sigmoid Function

3. NEURAL NETWORK LEARNING

a. *Learning Rules*

As has been mentioned previously, the purpose of the network is to take a set of inputs in the form of features represented as numbers in an input vector and map them to one in a category of probable output types, represented as the activation levels of the output neurons in an output vector. These output levels can take on any values in the set (0,1), with values near zero representing low activity levels and values near one corresponding to high activity levels for the associated neuron. For the network to do this it needs to have "learned" what the output categories are and what input vector

features are representative of a particular type of output vector. There are a number of clever and innovative ways of doing this [Ref. 3]. The method chosen for this work and that which will now be discussed is known as supervised learning utilizing the backpropagation algorithm which is based on the generalized delta rule.

Simply put, the goal is to present the network with exemplars of each type of input vector that it is expected to learn and then "tell" it that these input vectors correspond to a given output vector. A neural network unlike the human brain is simply computer code, thus the way it is "told" information is by way of numerical valued vector input. Numbers which represent features common to an output category type are presented to the network at the input layer. These numbers are then mapped through the network to the output by way of the transfer function operating on neurons and connections to arrive at final values at the output neurons. This process is then repeated a number of times for different exemplars of the various output vector types. During this "training" process the desired vector output is also provided to the network. An error is then calculated for the process. This error, in its simplest form compares the difference between the "perfect" or "desired" output activity for the given input, and the actual output neuron activation level calculated by the network. This error is then backpropagated through the network and it adjusts itself to minimize this

error. The manner in which the error is backpropagated and the way in which the network "adjusts" itself form the basis of the learning occurring in the network.

b. Generalized Delta Rule and Backpropagation

The final concepts which need clarification are the manner in which the network learns the associations necessary to perform its feature based recognition. As previously mentioned this is done by backpropagating the output error to the input and repeating the training presentation. Learning occurs in the form of adjustments of the weights representing the mathematical strength of connections between neurons. Through repeated presentations of the training vectors these weights are slowly adjusted to facilitate reduction in the output error. This is accomplished practically through use of the generalized delta learning rule to adjust the weights and the backpropagation algorithm to communicate the information back through the network.

(1) Generalized Delta Rule. The generalized delta learning rule states that the change in the weight of the connection between the i^{th} and j^{th} neurons is proportional to the difference between the error input to the i^{th} neuron and the activation of the j^{th} neuron or:

$$\Delta w_{ij}=\epsilon\delta_i a_j \tag{8}$$

Where

ϵ = a learning rate parameter which determines how fast the network changes the weights

$\delta_i=(t_i - a_i)(f_i)'(net_i)$ for an output neuron

t_i= The training input to the i^{th} neuron

a_j= the activation of the j^{th} input neuron

f'_i=Derivative of the activation function with respect to a change in the net input to the neuron

$net_i=\Sigma a_j w_{ij}$ + $bias_i$

The bias term mentioned above is the same as was described in association with the description of the connection geometries of Figure 3. The δ_i given above is for an output neuron. For the non-output neuron δ_i is given by:

$$\delta_i=(f_i)'net_i\sum_j \delta_j w_{ji} \tag{9}$$

It can be shown that this rule will find a set of weights that drives the error arbitrarily close to zero for every set of patterns in the training set if such a set of weights exist. Such a set of weights will exist if, for each input pattern target pair, the target can be predicted from a linear combination of the activation of the inputs. [Ref 4]

(2) Backpropagation. To complete the discussion of how this new information is communicated to the network a brief explanation of the backpropagation algorithm is presented. The basic idea of the backpropagation method is to

combine a nonlinear system capable of making decisions with the objective error function of Least Mean Squares and gradient descent. The objective error function for Least Mean Square error is:

$$E=\frac{1}{2}\sum_{i}(t_i-a_i)^2 \tag{10}$$

To implement this idea one must be able to compute the derivative of the error function with respect to any weight in the network and then change the weight according to the rule:

$$\Delta w_{ij}=-k\frac{\partial E}{\partial w_{ij}} \tag{11}$$

The "k" in Equation 11 above is just a proportionality constant.

The application of the back propagation rule, then involves two phases: During the first phase the input is presented and propagated forward through the network to compute the output value a_j for each neuron. This output is then compared with the target, resulting in a δ term for each output neuron. The second phase involves a backboard pass through the network (analogous to the initial forward pass) during which the δ term is computed for each neuron in the network. Once these two phases are complete, the weight error derivatives (Equation 11) can be used to compute the actual

weight changes on a pattern by pattern basis, or they may be accumulated over the entire ensemble of patterns. Additional details can be found in "Parallel Distributed Processing" from which the foregoing discussion was taken.[Ref. 4]

III. FEATURE BASED NEURAL NETWORK CLASSIFIER

As discussed in the introduction, the goal of this thesis is to demonstrate a feature based acoustic transient classifier. This section describes the design and operational details for the feature based classifier.

A number of design considerations and parameters play into the question of designing a neural network which can perform this type of classification task. These include:

1) Characterization of input data sets.

2) The type of network best suited to perform the classification task.

3) The size of the network needed to perform the task.

4) Decisions on training data and training time such that network performance is optimized.

Each of these will now be discussed in some detail as they relate to the classification task at hand.

A. INPUT DATA CHARACTERISTICS AND ANALYSIS

Data used in this thesis consists of raw times series voltage data for three different types of acoustic transients. For discussion purposes for the remainder of this thesis these transients will be referred to as type I, type II, and type III transients. These transients were recorded at sea in the presence of the type of background noise described in section I. In addition to the raw times series data another set of

signal data were produced by signal processing to extract relevant information features contained within an individual record or transient. Unclassified examples would be such things as frequency content, amplitude, density of the power spectrum etc. When necessary these features will be referred to as feature a,b,c, etc. All data were obtained from the Naval Surface Warfare Center (NSWC) and all data preprocessing was done there. These data were processed by NSWC to characterize each transient event in terms of 45 different features. Some of the features however provide redundant information so that the final processed data set used in this portion of the thesis utilized only 31 of the features.

The acoustic transient identification question is a matter of pattern recognition. In other words, one could ask if there is structure in transient type I which is different than type II, and III. Additionally one may ask are there features in exemplar #1 of type I which are similar to the features in all other type I transients. If this is the case then a neural network may be able to recognize and more importantly recall patterns in this structure and thus distinguish between class types. Further one hopes that there are unique features within a data class which clearly distinguish it from other data classes.

1. Euclidean Distance Analysis

To address these questions, related to classification, a substantial effort was made to characterize the data. With

data of this type (i.e. feature extracted) characterizing the input data by class is not a trivial question. One technique which was utilized in this research was to simply treat the input data as vectors arranged on a 31 dimensional hypersphere. This approach then allows the calculation of euclidian distance (D) on the hypersphere from the tip of one vector, say exemplar 1, to the tip of all other vectors in the space.

$$D=\sum_{i}\sum_{j}\sqrt{(x_1(j)-x_i(j))^2} \tag{12}$$

The following four figures, Figures 5 through 8, illustrate euclidean distance for vectors in the data set.

The first figure, Figure 5, represents the euclidean distance from a type I vector plotted against 150 vectors chosen at random and representing all data classes. The remaining three figures, Figures 6 through 8, represent one vector from each data type graphed as euclidean distance from the remaining vectors of its type in the data. Inspection of the graphs reveals considerable variability, especially in Figure 5, which represents all data types, indicating there are a number of different data classes within the entire data set. However, a closer look at Figures 6 through 8 show that the data can in fact be categorized into distinct classes. For example for the type I data of Figure 6 there exist 5 distinct groupings. The first grouping contains those 4 vectors with a

total distance less than 0.2 x 10^4, the next grouping occurs between 0.4 x 10^4 and 0.6 x 10^4, the largest group is a set of data centered near 0.95 x 10^4, a fourth group consists of those points with distances between 1.1 x 10^4 and 1.5 x 10^4, and finally the last group consists of those 6 vectors represented as the large spikes with distances exceeding 1.8 x 10^4.

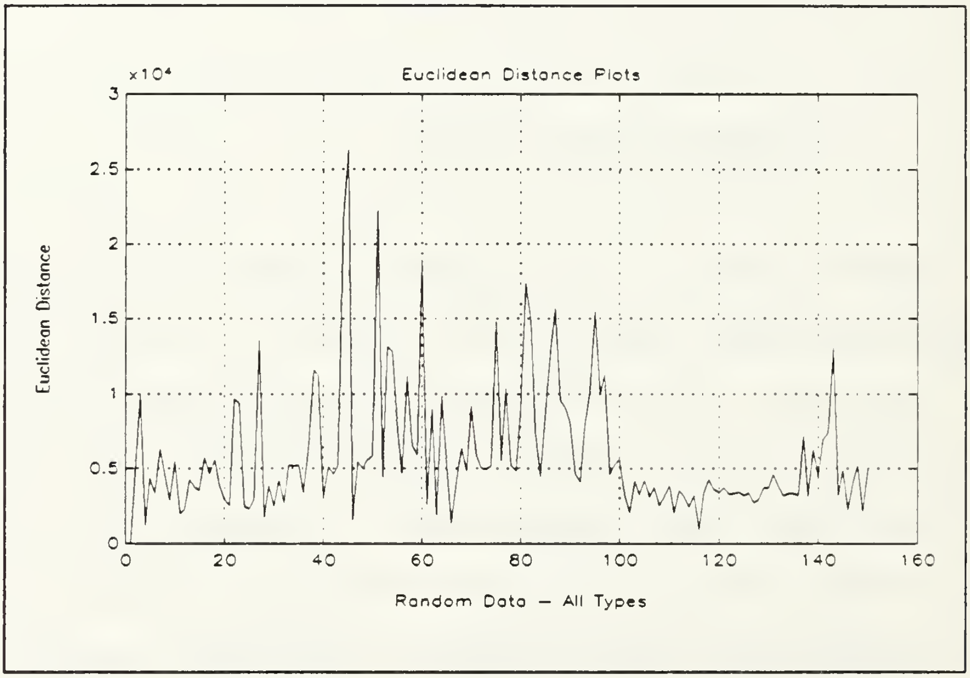

Figure 5: Euclidean Distance for all Data Types

This delineation is important because it points to the fact that the data can be characterized by a set of common features. Although only one vector has been chosen to illustrate the euclidean distance analysis, these vectors are representative of the data set and euclidean distance plots for other vectors in the data set provide the same analytical results.

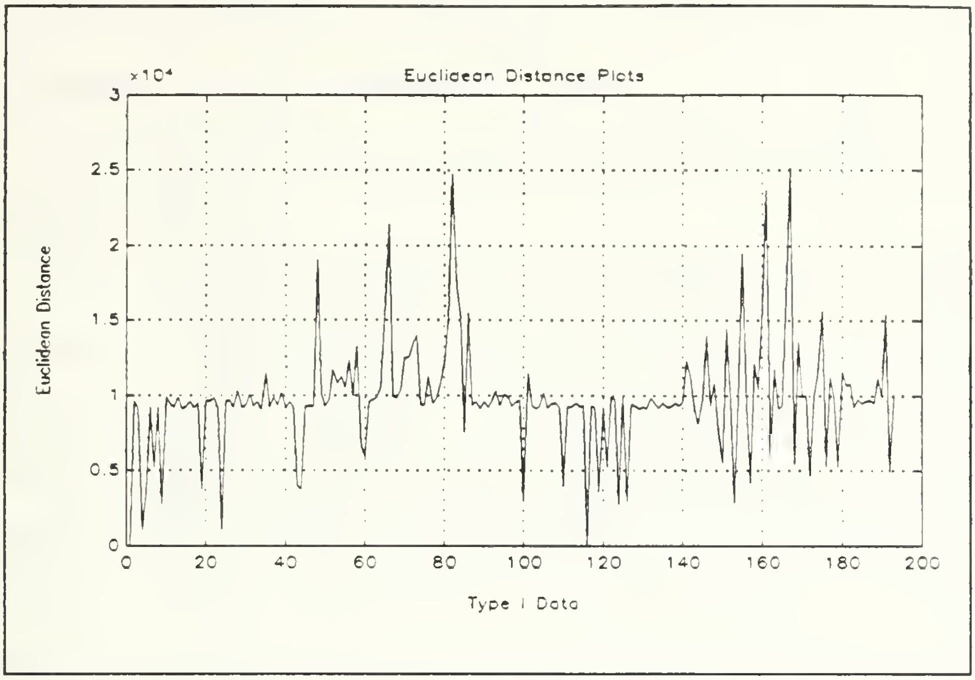

Figure 6: Euclidean Distance for Type I Data

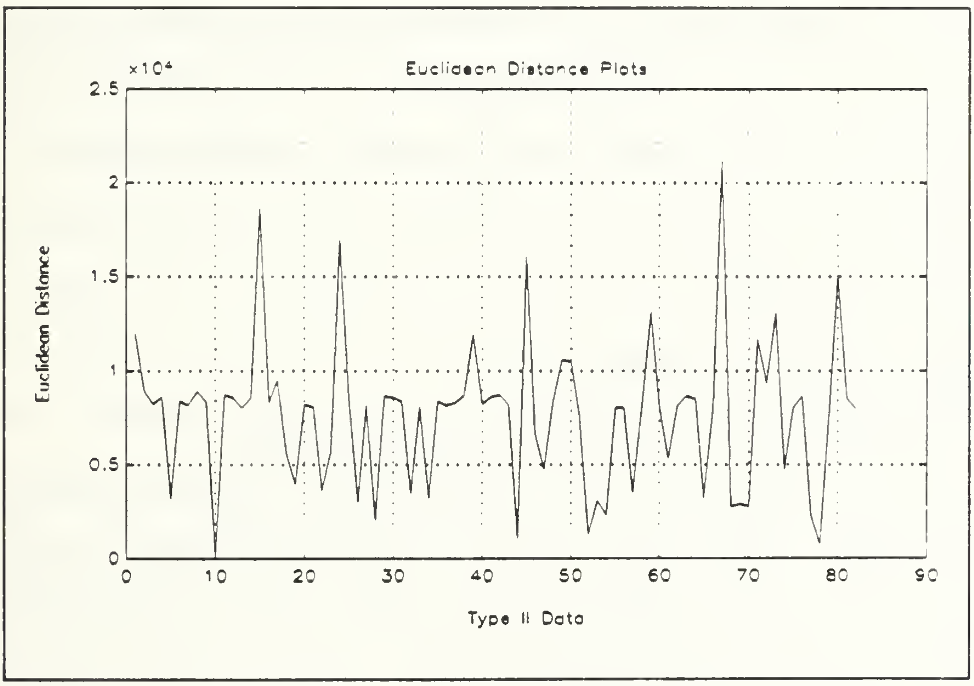

Figure 7: Euclidean Distance for Type II Data

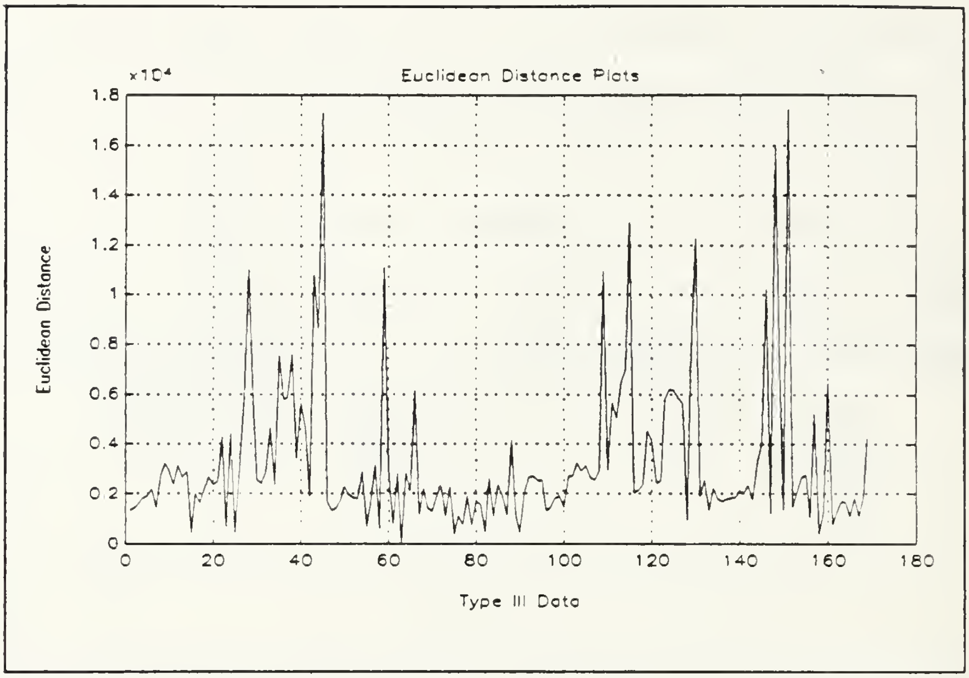

Figure 8: Euclidean Distance for Type III Data

Euclidean distance will be an important characteristic to consider when making up the final training and test data sets, as it is particularly important that all data subgroups within a given data type be represented in the training data set if the network is to perform recognition tasks on all of the test set satisfactorily.

B. NEURAL NETWORK CONSTRUCTION

1. Network Type and Size Considerations

The next step in the classification task was to settle on a network type. This is an important neural network question and will certainly differ from task to task. When answering this type of question there simply is no substitute for domain knowledge. Knowledge of the nature of acoustics and acoustic transients are the keys to making the correct choice.

This thesis utilized a heteroassociative backpropagation single hidden layer network to perform the classification task. This type of network is particularly suited to pattern recognition.[Ref. 5].

The next question which must be addressed is the size of the network which is best suited to perform the task. For this portion of the analysis the size of the input layer to the neural network is fixed by the number of individual parameters which are used to characterize each exemplar in the data set. The original data contained 45 individual parameters or features, 14 of which were redundant or were used for data tags rather than to convey signal information, thus the final data set contained 31 individual parameters characterizing the data into one of three types. This fixed the input data layer size at 31 neurons.

Next one must decide on the number of hidden layers and neurons which will enhance efficient and reliable network performance. Few theoretical studies are available to guide neural network practitioners in answering this important question. Neural Ware, Inc., a professional Neural Network Engineering Corporation does provide some guidance [Ref. 5]. Neural Ware suggests that the number of hidden layer neurons is proportional to the ratio of the number of exemplars in the data set to the sum of the nodes in the input and output layers:

$$H = \frac{d}{f(m+n)} \tag{13}$$

Where

d = # of exemplars in the data set

f = Arbitrary number between five and ten

m = # of neurons in the output layer

n = # of neurons in the input layer

For the work cited here this number computed to three neurons in the hidden layer. A three hidden neuron network was built and tested but performed poorly. This guidance may be useable for very large data sets but proved to be of little use in the construction of a hidden layer for the work considered here.

a. Singular Value Decomposition

Recall from section II that a neural network learns by adjusting connection weights between neurons. These weights are stored in a weight matrix and updated during the training process. This weight matrix is nothing more than an array of numbers and like any other numerical array is characterized by certain properties. One such property of importance when investigating the hidden layer size is the number of singular values in the weight matrix. The number of singular values in the weight matrix determines the number of linearly independent eigenvectors necessary to fully span the vector space. This number in turn provides a basis for the

number of independent features in the data and thus provides a good starting point for determining the number of neurons necessary in the hidden layer for network convergence.

The data considered here was analyzed and decomposed to singular values utilizing MATLAB, a commercially available signal processing tool. MATLAB code was written to capitalize on the resident singular value decomposition feature.

Figure 9 below represents the singular value decomposition of the data set.

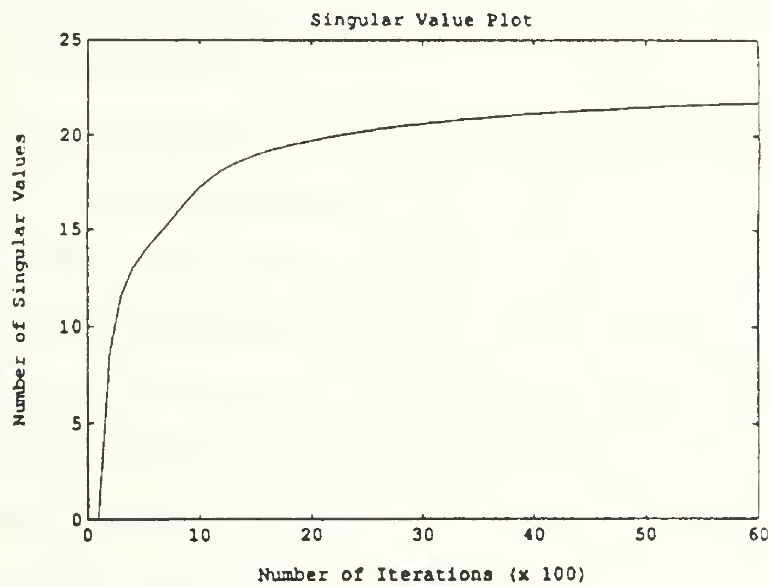

Figure 9: Singular Value Decomposition

Scrutiny of Figure 9 shows that the data contains approximately 21 singular values. This then forms a basis for

determining the number of individual independent elements in the data set that a hidden layer might be expected to extract. Note that the curve in Figure 9 continues to rise slowly even after 6000 iterations, indicating the presence of perhaps a few more singular values. The number of singular values extracted by the MATLAB software of course depends on an operator selectable threshold. Had a smaller threshold been used the number of values extracted would have been slightly higher.

Networks containing 21 neurons in a single hidden layer and networks which distributed the 21 neurons between two hidden layers were built and tested. Results are reported below.

Theoretical discussions of this subject suggest experimenting until satisfactory performance is achieved. Using the singular value decomposition above as a guide, experimentation was conducted which attempted to find the best number of hidden layer neurons.

This experimentation led to a final network size of 31 input neurons, 25 hidden neurons in a single layer, and 3 output neurons. This network was built, tested ,and found to be efficient and reliable. Results of the performance of this network are discussed in the results portion of this section.

C. TRAINING THE NEURAL NETWORK CLASSIFIER

Often an important consideration in neural network training and performance is the content of the training file

relative to the test file and the length of training time required to ensure satisfactory network performance. These issues will now be addressed.

The fundamental performance test that a neural network must pass is an ability to learn and then recall the entire data set. This is important because failure of the network to be able to do this may point to inconsistent or mislabeled data, the wrong type of network for the task, or simply a problem which is not suitable for a neural network to solve. The network described above satisfactorily learned and recalled the 458 exemplar data set to 100% accuracy. This being achieved it was necessary to break the data set up into training and test sets.

The first data split consisted of placing the first half of the 458 exemplars in a training file and the second half of the 458 files in a test file. Performance for the network trained on the first 229 exemplars and tested on the last 229 exemplars was satisfactory but not optimum. Results of this testing is discussed below and compared to other networks in Table 2.

The next step in training and test set construction was to split the data in half by random selection, hoping that enough exemplars of all data classes within a type would exist in both sets to allow for satisfactory performance. This delineation did in fact result in better performance. The network still however was unable to recognize a small

percentage of all data types. Further, these results led to questions concerning characterization of the data set within exemplar types. This question was for the most part resolved by the use of Euclidean distance as a class indicator. Having determined, through this analysis, that many different data classes existed within a given data type, the question still remained as to whether enough unique features existed to allow a neural network to separate data by type during training and recognition.

Individual misclassifications were then examined and a few more exemplars of odd or infrequent data classes were moved from the training set and placed into the test set, and the network was again tested. This network performed quite well, and its performance along with a comparison of results obtained from the other networks mentioned above are discussed in the results portion of this section.

Finally the last consideration relative to network training was to find the training time, which resulted in optimum network performance, characterized by the fewest number of misclassifications in the shortest possible training time. A procedure similar to that followed by Hecht-Nielsen was utilized to address this performance issue [Ref. 3]. Figure 10 below shows network performance graphed as the number of misclassifications versus the number of training cycles.

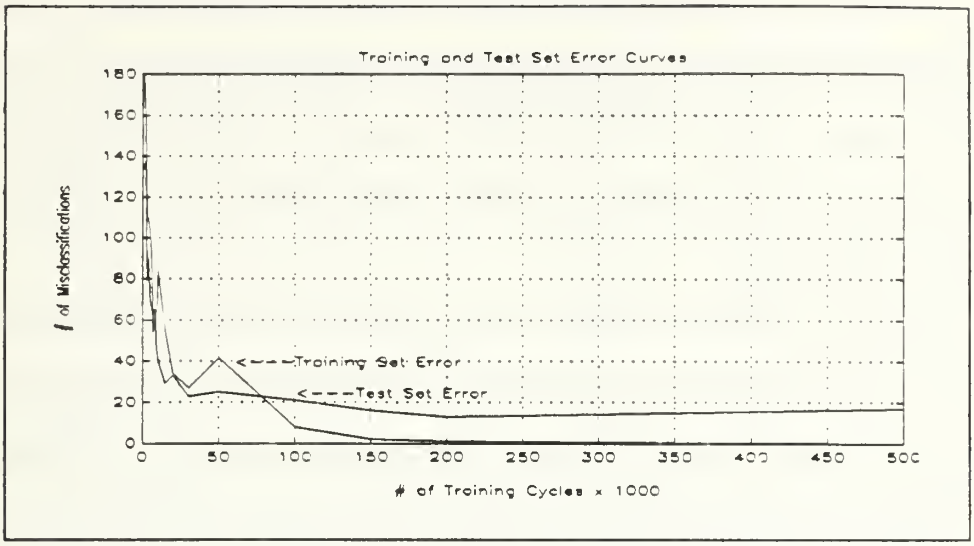

Figure 10: Optimum Network Training Time

In training a neural network, the network is first trained on and then subsequently tested on the training set. This demonstrates that the network is suitable for the task at hand. During this type of training the recognition error should continue to decrease indefinitely. However when training on the training set and then testing on the test data one finds that the error will eventually reach a minimum , and then begin to increase again as the network simply begins "memorizing" the input data set. It is this minimum in the test set curve which represents the point of optimum training time. As seen from the Figure 10 this occurred for this network at approximately 220,000 cycles of training.

D. RESULTS: TESTING THE FEATURE BASED NETWORK

A number of different networks have been described in this section. Comparative results for four of these networks is now presented in tabular form. These networks are:

1) A 31x25x3 network which was trained on 50/50 data split with the data being selected sequentially.

2) A 31x25x3 network which was trained on the data split 50/50 again but this time the data split was made by random selection.

3) A 31x21x3 network which was trained on the final data split. This data split consisted of a 50/50 training/test split in data, with the data being selected at random. After the random data selection, Euclidean class analysis was done on both sets and some additional exemplars were moved from the test to the training set to ensure all classes of data were included in the training data set.

4) A 31x25x3 network trained on the final data set, i.e. the same data set used in network #3.

Before presentation of results it should be noted that each network was trained to the same standard. This was done by training Network 4 to the optimum point as discussed in the network training section above, and noting the rms error for the output neurons. Networks 1 and 2, being the same size, were then trained to the same number of cycles. Network 3 being smaller in size was trained to the same rms error.

The final data set for the best network (network # 4) consisted of the data breakdown shown in Table 1, and the testing results for these feature based networks are summarized in Table 2 below.

TABLE 1: DATA BREAKDOWN BY TYPE IN FINAL DATA SET

# of Exemplars by Data Type	Training Set	Test Set
Data Type I	115	86
Data Type II	54	33
Data Type III	90	80

TABLE 2: RESULTS FOR FOUR FEATURE BASED NEURAL NETWORKS

Recognition percentages	Type I Data (#correct/86)	Type II Data (#correct/33)	Type III Data (#correct/80)
Network # 1 (31x25x3) Seq. Data	0.26 (22/86)	0.55 (18/33)	0.53 (42/80)
Network # 2 (31x25x3) Random Data	0.89 (76/86)	0.87 (29/33)	0.92 (74/80)
Network # 3 (31x21x3) Final Data	0.71 (61/86)	0.71 (23/33)	0.96 (77/80)
Network # 4 (31x25x3) Final Data	0.92 (79/86)	0.94 (31/33)	0.95 (76/80)

Some analysis of these results is now in order. Comparison of rows three and four shows improvement when training on the same data set with a network which contains 25 vice 21 neurons in the hidden layer. This is evident by comparing the improved recognition percentages in row four (.92 for type I data) over those in row three (.71 for type I data). This suggests that there are more than 21 independent features in the data which the network is using to fully characterize and classify the data.

Recall that singular value analysis indicated that the number of units in the hidden layer should be of the order of 21. Good performance was obtained with a network of 25 hidden units.

Next compare rows one and two of Table 2. Here we see quantitatively the importance of random data selection in data enhancement. Compare the improved recognition percentages in row two (0.89 for type I data), where data was selected randomly to form training and test sets, to that in row one (0.26 for type I data), where data was formed by splitting the whole data set in half sequentially. Random selection clearly improves the likelihood of including all data classes within a data type.

Last consider rows two and four. The 3% improvement shown in the recognition percentages of network four (0.92 for type I data) over network two (0.89 for type I data) is a direct result of the euclidean distance analysis on data class

structure. This improvement was realized by using euclidean distance to ensure that exemplars of all data sub classes within a type were included in the training set.

The implications of the success of network four and a comparison with other networks considered in this thesis are discussed at length in the final section of this thesis.

IV. TIME AND FREQUENCY DOMAIN NEURAL NETWORK CLASSIFIERS

Having considered the detection of short duration acoustic transients by neural computing methods in "feature space" it is instructive for comparative purposes to consider detection of these transients in the time and frequency domains.

A. TIME DOMAIN NEURAL NETWORK CLASSIFIER

Recall that the original data for this thesis was obtained by recording the analog voltages in a continuous time series from a waterborne buoy. This data was then sampled at a fixed sampling rate (i.e. digitized). The acoustic transients were then electronically "snipped" from the digital recording and processed to parameterize them into 31 distinct features. This section of the thesis considers the detection and classification of the original digitized time series data.

1. Time Domain Data Analysis

Each snipped times series contains within it the acoustic transient of interest. See Figure 11 on the following page. Figure 11 is a typical type I transient time series record. It consists of 3000 points of raw data representing one acoustic transient and the background noise which surrounds it. As is clearly evident from Figure 11 most of the information content in the record consists of mere background noise. It is neither necessary nor desirable to present the majority of this background noise to a neural network.

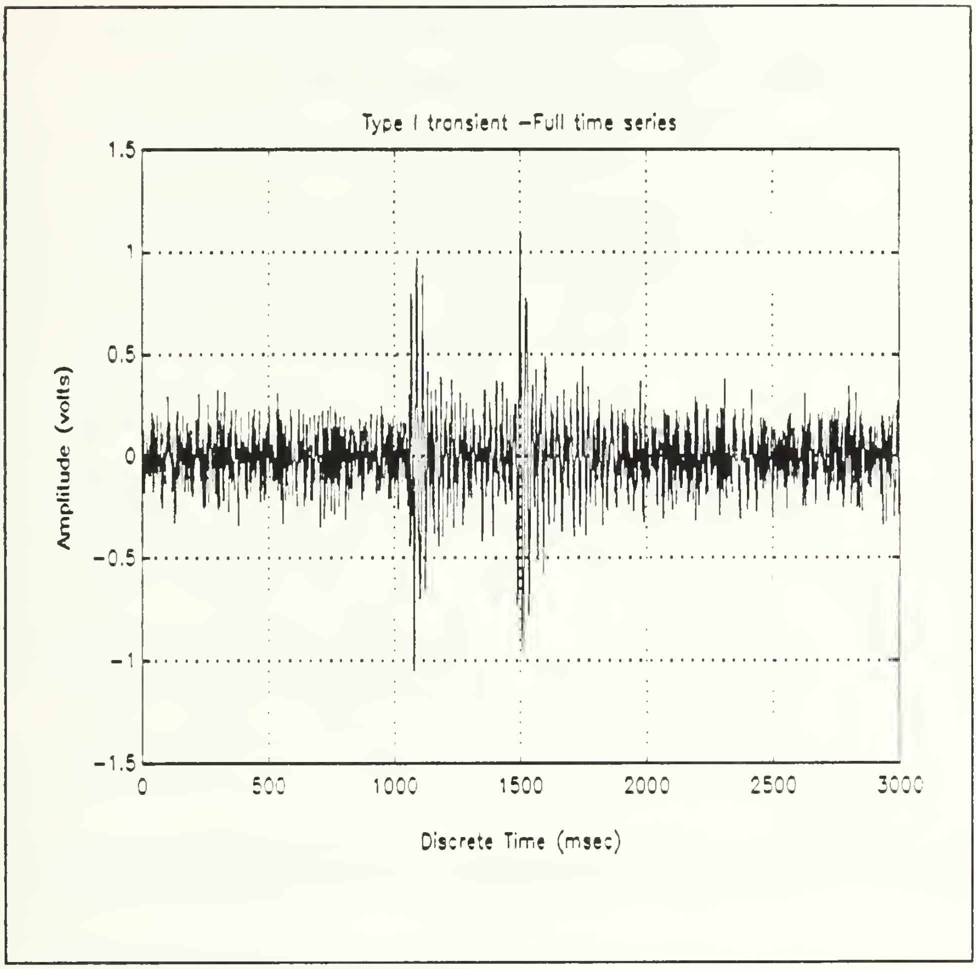

Figure 11: Type I Transient; Full Time Series

One significant disadvantage of doing so is that background noise is common to all transient types and thus provides no new information to the network by which it can make discrimination in the classification process. Additionally the length of the record determines the number of input neurons to the neural network. Network size and more importantly training time is significantly reduced by removal of this noise.

Figure 12 below is the same type I transient (The 2nd peak in Figure 11). In Figure 12 just the 150 points on either side of the transient peak has been retained.

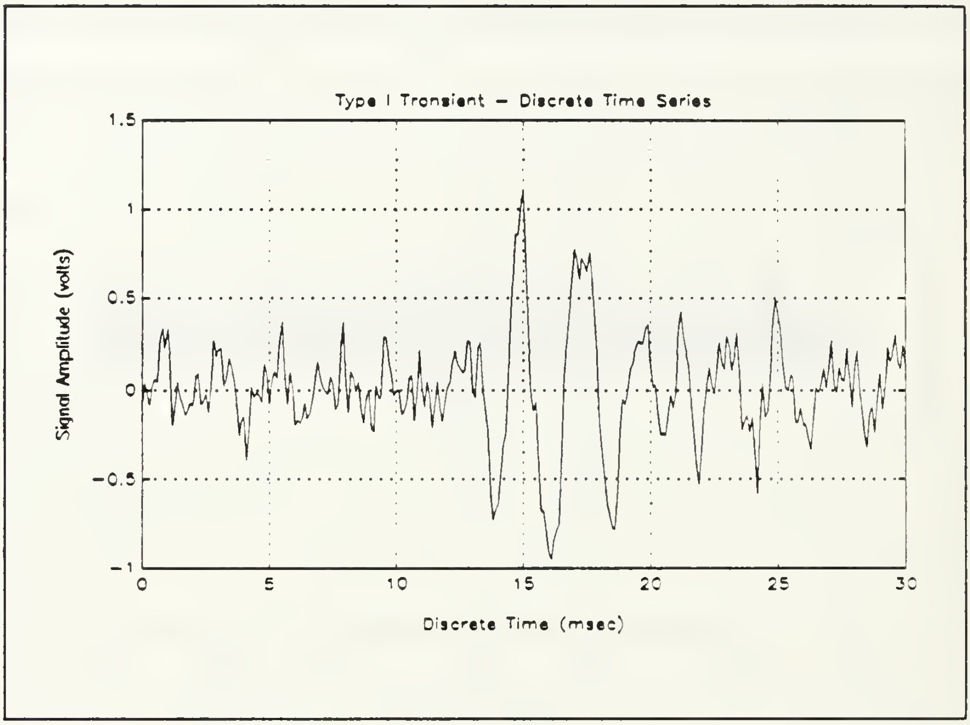

Figure 12: Type I Transient; Discrete Time Series

This representation of the data retains the essential information relevant to classification of the transient but is much reduced in size, and thus will allow a neural network classifier which can be trained in fractions of the time to train on the full record. Figures 13, and 14,on the following page present type II,and type III transients for comparative purposes. Close inspection of Figures 13 and 14 when compared to Figure 12 yields subtle but important differences in the structure of the signals.

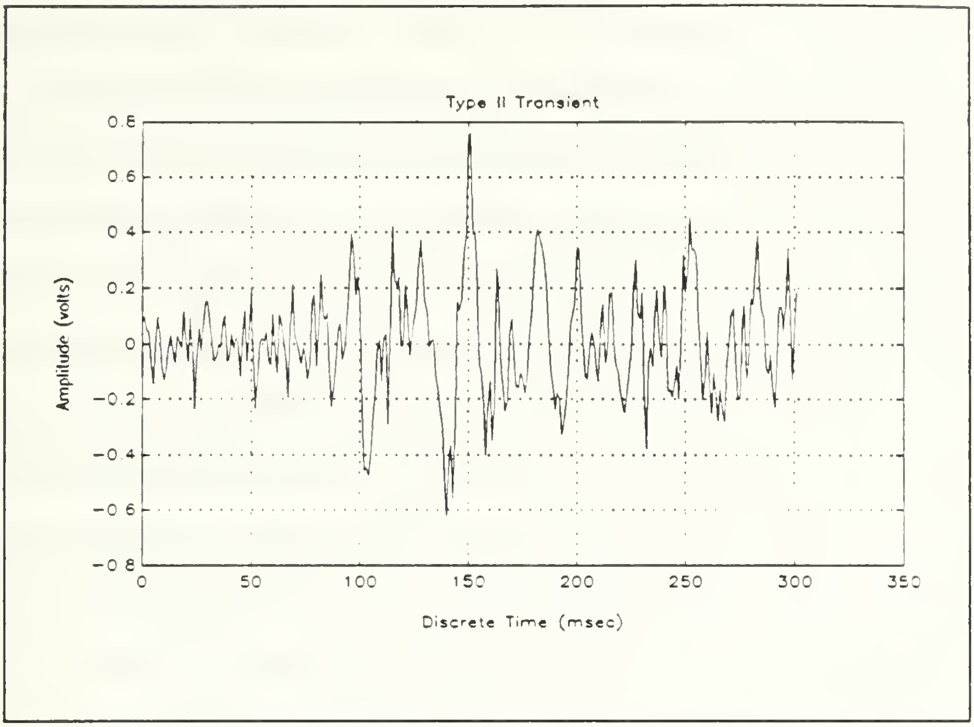

Figure 13: Type II Transient; Discrete Time Series

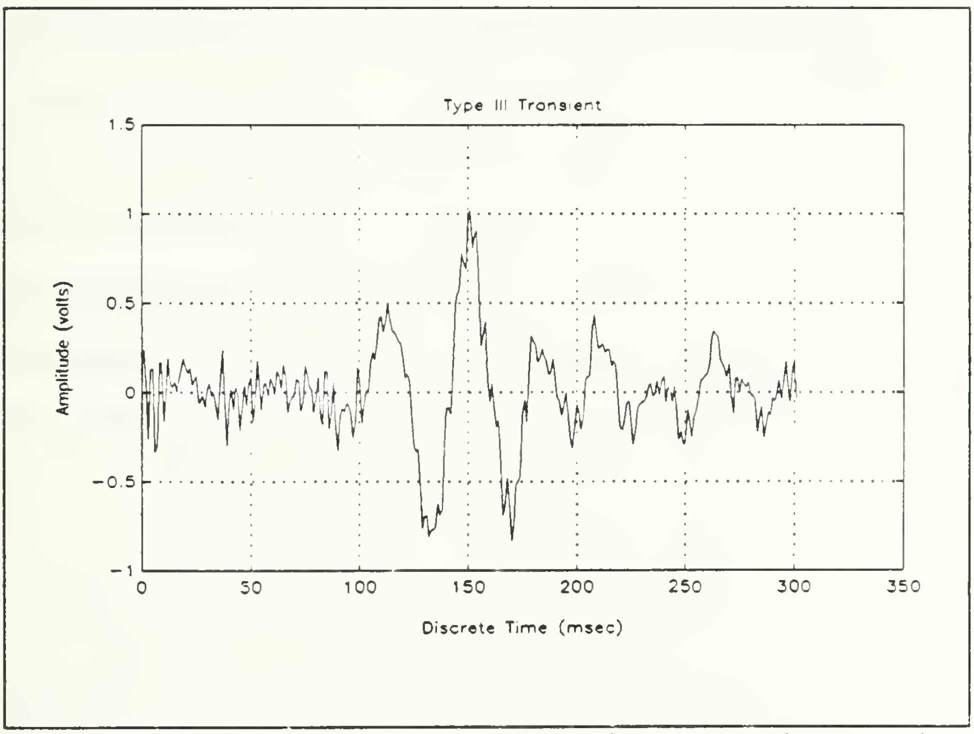

Figure 14: Type III Transient; Discrete Time Series

These differences are more marked in the frequency domain and will be discussed in detail later. However note that the type I transient shows a distinct and sharp rise followed by a steady decay, which is characteristic of an exponentially damped decay. Compare this to the type II and type III transient which show more gradual rises. These latter type of transients seem to more slowly build to peak values and then slowly decay as opposed to a sharp burst of energy which then decays characteristic of the type I transient. It is features such as these that the neural network will use to distinguish between the types of transients.

2. Training and Test Set Data Construction

a. Training The Network

Next it is relevant to consider the distribution of the training and test data sets. A detailed discussion of how data can in general be split was covered in section III. In section III recall that the final data set was split into a training set consisting of 259 exemplars and a test set of 199 exemplars. NSWC graciously provided at the authors request all 458 of the feature based exemplars and 60 exemplars of times series data. The 60 time series data exemplars (Figure 11 represents one such exemplar) represent the time series from which 60 of the 458 exemplars of feature based data were extracted. Thus as performance comparison of neural networks in feature space, the time domain, and the frequency domain was a stated goal of this thesis, training and test data sets

in the time and frequency domains were split to ensure that their feature based counterpart remained in the same data set, either training or test. That is if a data vector was in the feature based training set and it was one of the 458 vectors for which time series data existed then its time series data also went in the time series training set, and likewise for data in the test set. As the training data set in feature space was larger than the test data set this led to a somewhat disproportionately large training data set in the time domain as well. One vector had to be eliminated from the time series data set leaving the remaining 59 vectors in the time domain to be distributed as follows:

TABLE 3: TIME SERIES DATA BREAKDOWN

# of Exemplars	Training Set	Test Set
Type I Exemplars	24	15
Type II Exemplars	5	4
Type III Exemplars	7	4

b. Results: Testing the Time Domain Network

Several networks were built and tested on the time domain data. All performed poorly. The network showing the highest success was a backpropagation multi layer network with 300 neurons in the input layer, 150 in the first hidden layer, 20 neurons in a second hidden layer and finally 3 output

neurons. This network was only able to correctly classify 60% of type I transients, 45% of type II transients and none of the type III transients. Although disappointing in performance this network did lead to some understanding of the factors which may make detection and classification tasks difficult for a neural network. Others studying this problem, i.e. transient pattern recognition in the time domain using real world data, have had trouble with consistently good recognition [Ref. 1]. The reasons for some of these difficulties will now be discussed.

3. THE ARTIFICIAL TIME DOMAIN NETWORK

In investigating the difficulties associated with this classification task, one has to first answer the question: "Is this task suitable for neural networks?". In the present case this translates to:" Can a neural network learn acoustic transient patterns in the time domain?".

In contrast to the problems mentioned above some researchers have studied this problem and produced excellent results [Ref. 6][Ref. 7]. To help answer the question in the preceding paragraph and to sort out why one task is achievable while the another is often not, artificial acoustic transients were built to serve as test and training vectors which could be easily manipulated for investigative purposes.

a. Construction of the Data Set

Figure 15 below shows an artificial transient generated for use in the following investigation. Figure 15 is

labeled as a type I transient. It was built with the original actual type I transient serving as a model, and comparison between the two shows some similarity. Comparison with Figure 12 reveals that both transients are preceded by background noise, and then jump suddenly to a peak value and then decay exponentially. Both show randomness but also some periodicity. Figures 16 and 17 below are exemplars of the artificial type II and type III transients. These also show some similarity to their real counterparts, as they were built with a build and decay vice burst and decay structure in mind, and are clearly distinct from one another.

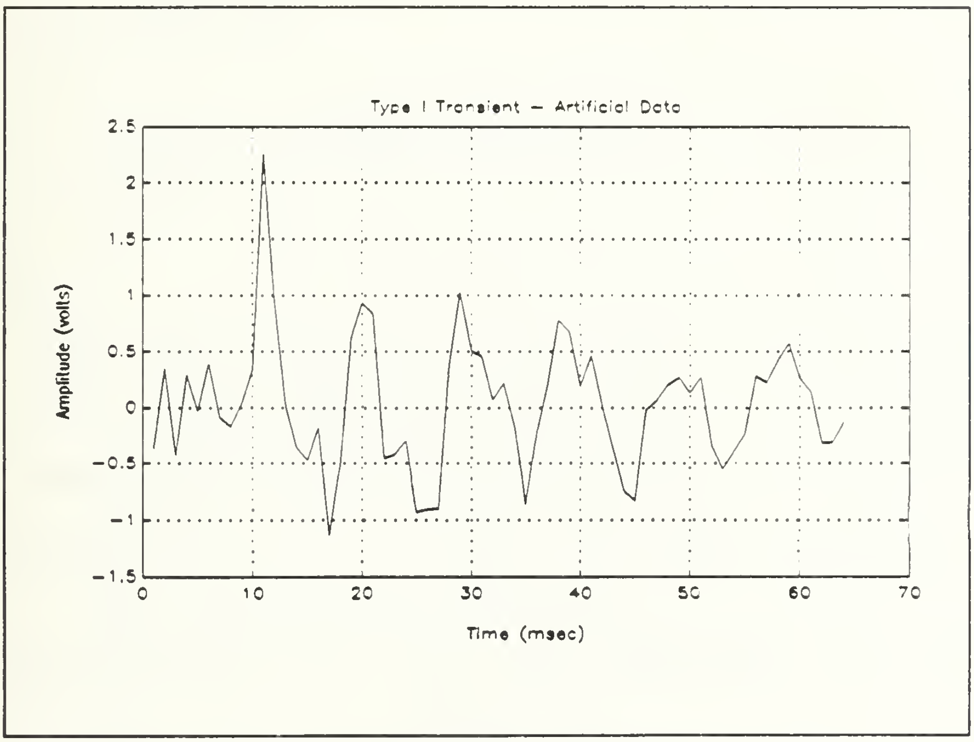

Figure 15: Type I Transient; Artificial Data

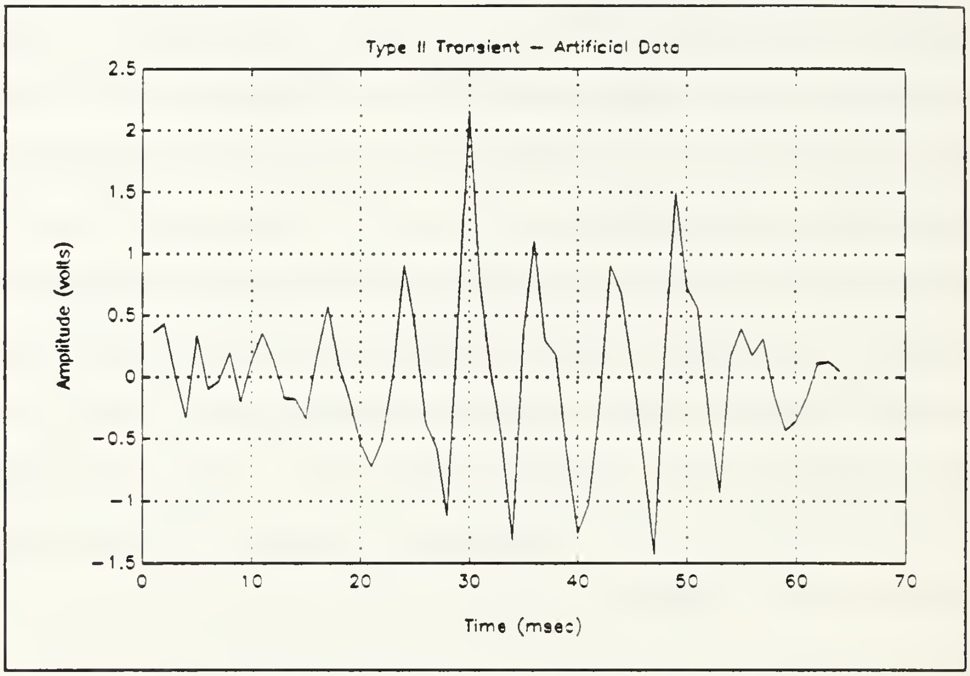

Figure 16: Type II Transient; Artificial Data

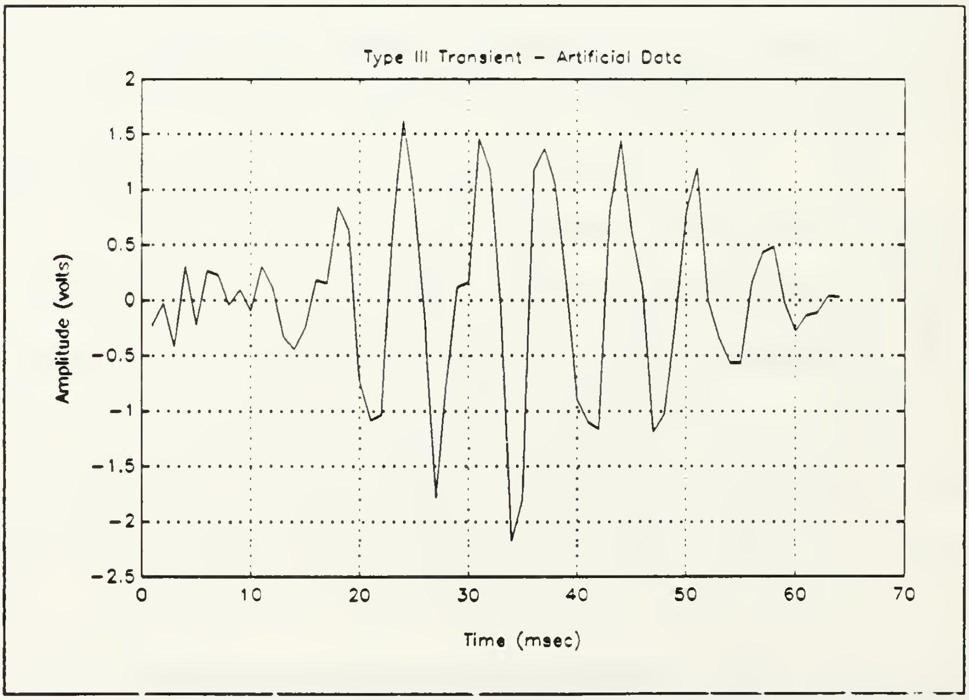

Figure 17: Type III Transient; Artificial Data

Regardless of the similarities between the artificially generated transients and their real counterparts there are some very important differences which are instructive to look at as they shed some light into why this task is so difficult in the time domain and point to some areas which may show promise for improvement.

In discussing these differences it is instructive to look at how the artificial transients were generated. The artificial transients were generated by consecutively adding together sine waves of 5 different frequencies, each with variable amplitude

Individual records were built in MATLAB from an equation of the form:

$$t_{ij}=\sum_{i=1}^{5}\sum_{j=11}^{64}(A_{ij}+bias_1)\sin(f_{ij}+bias_2)\,e^{-aj} \tag{14}$$

Where

t_{ij}= Transient voltage

A_{ij}= Initial transient amplitude

f_{ij}= Frequency of the transient component

bias1=Random bias term put on each point to produce minor statistical fluctuation.

bias2=Random bias term put on each frequency to produce minor frequency instabilities

a= Decay constant for exponential decay of signal

As one can see from Equation 14 the signal vector starts at point 11 and runs to point 54, generating a 54 point long vector. Each vector is preceded by 10 points of random noise, to give a total vector length of 64 points. A 64 point long vector was chosen to enhance transformation into the frequency domain if desired. 100 exemplars of each of the three types of transients were built and then the data was split in half to form training and test sets. Figure 18 below shows all 50 of the type I transients plotted together. This figure is included to give the reader a sense of the variability in this data even though it has been artificially generated.

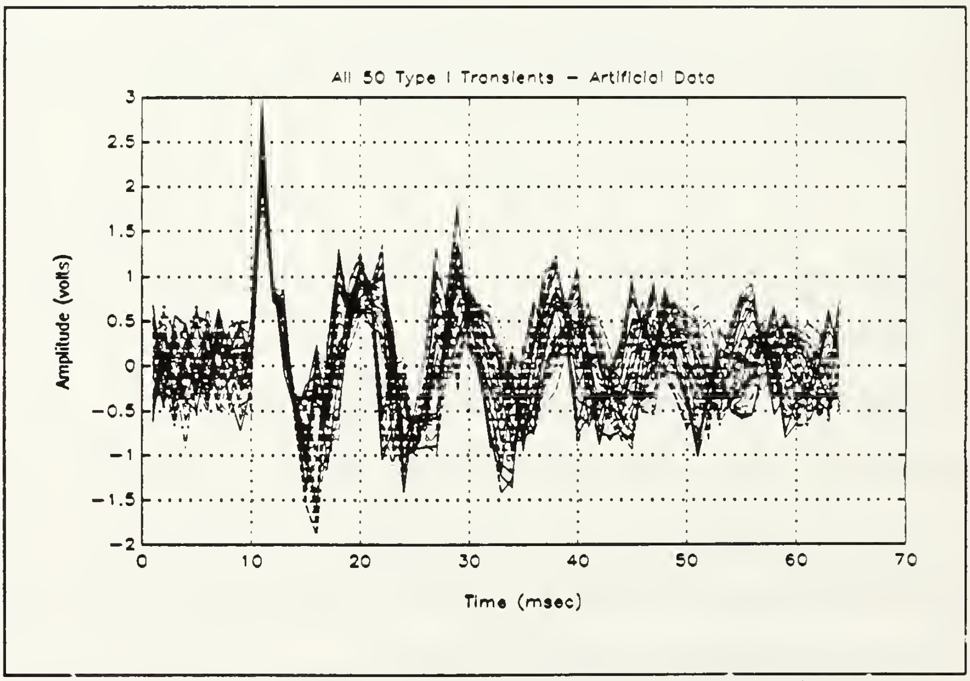

Figure 18: All Type I Transients; Artificial Data

b. Results: Testing on Artificial Data

A backpropagation multi layer network with 64 neurons in the input layer, 20 neurons in the first hidden layer, 10 neurons in the second hidden layer, and 3 output neurons was built and tested. Performance results were excellent with the network recognizing 100% of the type I and type II transients, and 94% of the type III transients. These performance statistics partly answer the fundamental question: "Can a neural network recognize and classify acoustic transients in the time domain?". It is now important to consider why the artificial network performance was so superior to the real data network performance.

B. COMPARISON OF ACTUAL AND ARTIFICIAL RESULTS

First consider the manner in which the real network data was split. This data was split by patterns in "feature" space. Patterns which characterize data as unique in one "space" may not be sufficient to uniquely separate data into the same distinct patterns in another "space". In this case splitting the data to preserve uniqueness in feature space apparently led to a training set in the time domain which did not contain exemplars of every data type.

Next, performance may have been degraded by the fact that few real world exemplars exist. A neural network is often a preferred pattern classifier because it has the ability to learn and generalize, however for the network to properly generalize it must see sufficiently many exemplars with

sufficiently many distributed features to make general observations about the data set. It is not likely that a network can do this with only 5 or 7 exemplars to train on, when each exemplar contains 10 or more features that the network is trying to use to make those generalizations.

Last consider the differences in the data itself. A careful review of the artificial data will show that :

1) There exists no noise in the signal portion of the data. This is not to say that there is no variability but rather that there is no noise in the signal of the same type which precedes the signal.

2) All artificial transients start exactly at point # 11.

3) All artificial transient signals are exactly 54 points long. Because of decay some of the signals appear to be reduced to the pre-signal noise level, but for the most part some signal still exists for all 54 points

4) All of the artificial transients are basically the same shape, where they differ results from statistical fluctuations.

All of the above items can be modified. For example random pink noise (similar to sea noise) can be added to the artificial transient signals. The signal start point can be modified etc. However one finds successive degradation in the networks ability to classify when these modifications occur.

As an illustration, artificial white noise (gaussian with mean 0 and standard deviation 0.5) was added to the artificial

data and the artificial network was again trained to an rms error of 0.01 and retested. The results were 98% recognition for type I transients, 70% recognition for type II transients and, 72% for type III transients. These numbers clearly represent a reduction in the networks ability to classify properly as might be expected, however recognition of type I transients remains quite good. Figure 19 below is a plot of the 50 type I vectors in this new data set. Compare these to Figure 18 which is the same data set without noise in the signal. Although Figure 19 is significantly more garbled, the dominant feature occurs "early" in the signal and thus tends to not be washed out as much as features occurring later in the signal. This is because of the small randomness in the length of the signals causing later features to overlap one another.

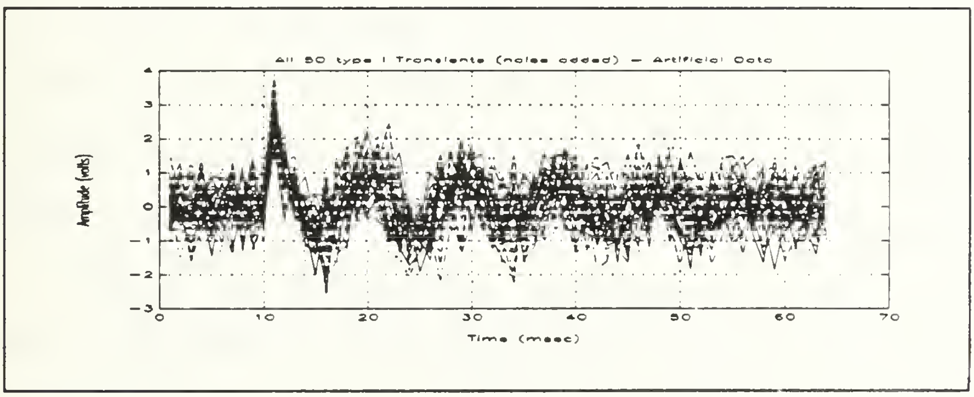

Figure 19: All Type I Transients with Noise Added

This aspect of the transient allows type I recognition percentages to exceed those of the other types. As the real data served as prototypical examples for construction of the

artificial data, one might expect better recognition of the type I real exemplars. This is in fact the case, partly because there are simply more exemplars than the other types and partly because the nature of the type I transient (burst and decay vice build and decay) lends itself to this self-preservation quality in the presence of noise.

The point of this analysis is that to enhance network performance care must be exercised with the manner in which the data is collected. Specifically if noise can be filtered during collection without suffering appreciable loss of signal this should be done (this turns out to be not practical for the real data set, see frequency domain analysis below). With respect to items two and three above it is important to pre-process data such that the data is "centered" in some fashion as it is presented to the network. This will of course depend on the nature of the data. For example one might want to ensure that the point of maximum amplitude occurs at the same input neuron, or that the signal always starts on neuron 10, etc. These are difficult questions to answer for data which contains signals of different lengths and amplitudes.

One of the reasons one might want to consider a neural network over other classifiers is its ability to generalize and thus overcome this problem of statistical shape fluctuation. We want and expect it to, for example, classify all "coins" as money or different types of "watercraft" as "ships". And indeed these networks are able to perform such

tasks if sufficient data exists to make these extended generalizations.

Of all of the conclusions drawn here the reader should be left with the sense that the primary reason that the time domain networks performed so poorly was because in the case of the real data network there was simply insufficient data, given the complexity of the individual vectors, to make the required generalizations.

C. FREQUENCY DOMAIN NEURAL NETWORK CLASSIFIER

Next, the original 59 times series exemplars were transformed to the frequency domain for analysis. Transformation to the frequency domain was accomplished by FFT. After frequency transformation a power spectral density of the form:

$$PSD(k) = |X(k)|^2 \frac{T}{N} \tag{15}$$

Where:

k = Discrete Frequency

X(k) = Fourier Transform Coefficients

T = Inverse of the sampling rate

N = Record length

was calculated and this data was used as the input data to the neural network.

Translation to the frequency domain has several inherent advantages over raw processing in the time domain. These are

discussed in detail in the section following this one. The only significant disadvantage of this transformation is the time required to pre-process the data.

1. Frequency Domain Data Analysis

As mentioned transformation to the frequency domain has several distinct advantages. These advantages and the role they play in the signal processing considered here are now discussed.

First, the size of the network required is automatically reduced to half of that required in the time domain. Figure 12 above shows one single time record which is 300 points long. When the FFT of this signal is taken a 300 point signal in the frequency domain is the result, however the signal is symmetric about the mid point, and thus the last half of the signal can be discarded. This results in a signal that is now 150 points long.

Next, all of the signals frequencies occur at the same neural network input neuron. To explain, if the signal contains 150 points and spans a frequency range of 0-4500 Hz then each point in the signal corresponds to an additional 30 Hz, making, for example, the 300 hz point always occur at input neuron #11. This "alignment of the signal" can be a significant performance barrier in the time domain as discussed above. Related to this is the fact that every signal can be of the same length regardless of the length of the transient in the original time record. The FFT will still

produce a 0-4500 Hz spectrum for example from a 300 point time record if the actual transient is only 100 of the 300 points or consumes all of the original 300 points. This has the effect of taking two transients which "appear" very much different in the time domain (because one is simply shorter) and producing equal representations in the frequency domain. The effect of this in terms of neural network recognition is to greatly simplify the classification task.

Last, it is sometimes possible in the frequency domain to "grow" the data set. If the original time record contains sufficiently long exemplars of the transient information then several cycles of the fundamental frequencies which characterize the signal should be present. This being the case one can sometimes split the record in half and FFT both halves of the time record to essentially produce two exemplars in the frequency domain from a single time domain record. Of course some information in the form of frequency resolution is lost as each frequency sequence is only half as long as the original and has only half of the resolution. Additionally one must exercise care when doing this to ensure that the first and second half of the time record are sufficiently similar to be able to perform this type of data multiplication. In the case of transient analysis this is often not the case, because the manner in which a transient begins or ends are significant in the characterization of the transient.

Another scheme which can sometimes be used in the case

of the signal asymmetry mentioned above is to take every other point from the time record and place it in a separate file. This has two effects, again the FFT length of the two new samples will be half of the original (giving up resolution but not bandwidth), and further it causes an effective halving of the sample rate which affects the bandwidth in the frequency spectrum. If the original frequency spectrum was 0-4500 Hz, the new signals will now only contain frequencies 0-2250 Hz. This may or may not be a problem for the classification task, depending on the frequency content of the original signals, but this method does not suffer from loss of the transient start information or transient termination information as the previously discussed method of data multiplication assuredly does. These methods have been discussed to serve as starting points for obtaining more data without field sampling should too little exist to reliably assess network performance.

Figure 20 below is the FFT representation of Figure 12 above. Several aspects of this signal are significant to the data preparation and presentation to a neural network.

Review of Figure 20 reveals that virtually the entire signal is contained within frequencies less than 1500 hz, the single exception being a very small component at 3063 Hz. Clearly the strength of this signal lies in the band 300-700 Hz with the dominant peak occurring at 499.5 Hz. Unfortunately this frequency band also contains the majority of noise from the ambient sea state [Ref. 8].

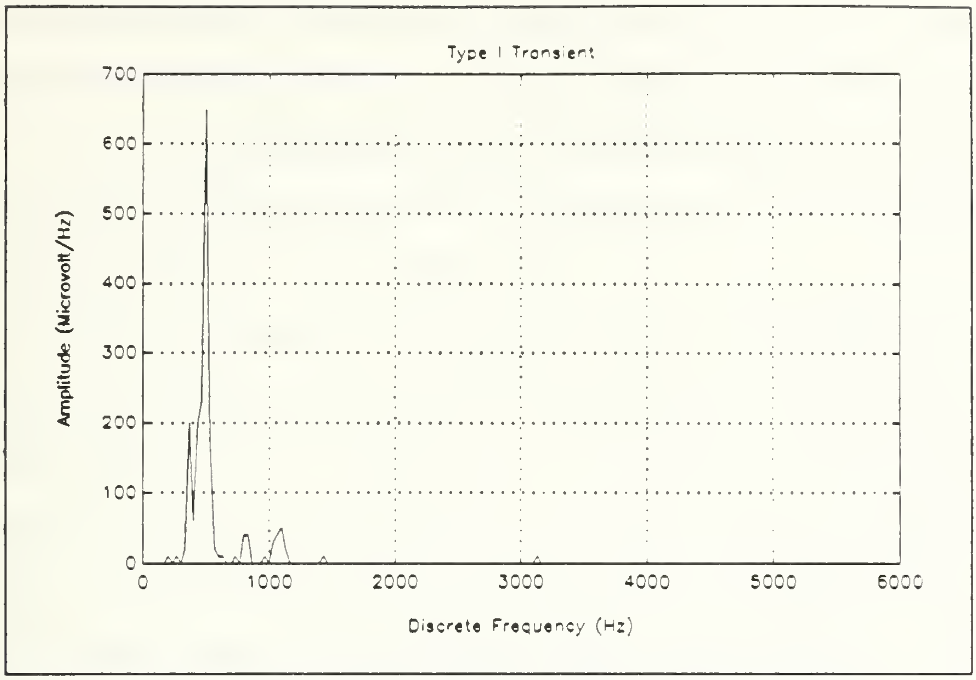

Figure 20: Type I Transient; Frequency Domain

Recall that one conclusion of the time domain analysis was that enhancement of the time domain signal could be accomplished through filtering the ambient sea noise, Figure 20 demonstrates this to be impractical for this data set. Last take note of the two smaller peaks centered near 800 Hz and 1100 Hz. Although these latter two peaks clearly are of less magnitude than the 499.5 Hz peak they are significant because they are pure signal and are sufficiently separated from the dominant ambient noise spectrum to serve as enhancing classification clues. Figures 21 and 22 below provide the frequency spectrums of type II and III transients for comparison. Comparison of Figures 21 and 22 with Figure 20 reveals many differences and a few similarities. First notice that dominant and secondary amplitude peaks are shifted in

frequency. Also note the grossly different amplitude scales (0-700 Microvolt/Hz for Figure 20, 0-5000 Microvolt/Hz for Figure 21, and 0-180 Microvolt/Hz for Figure 22).

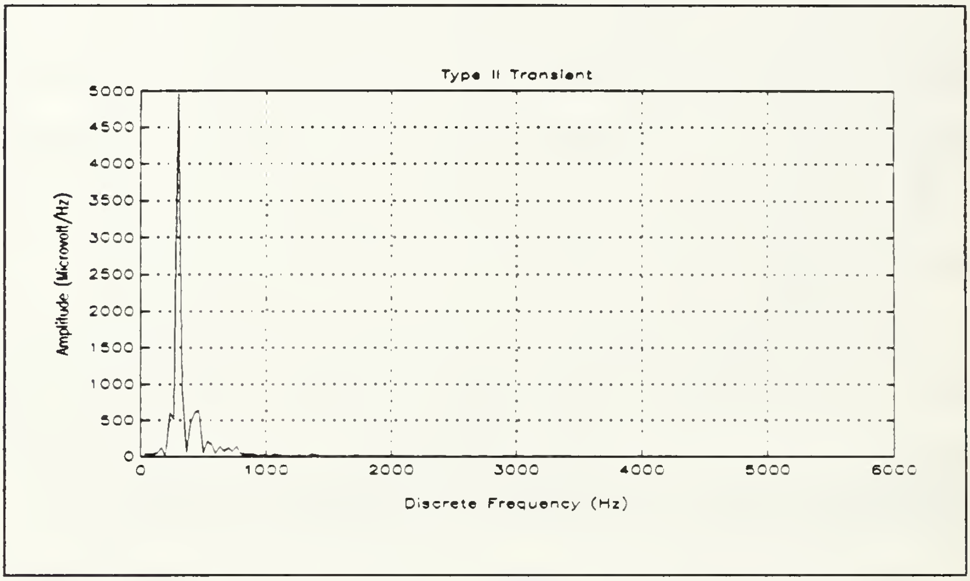

Figure 21: Type II Transient; Frequency Domain

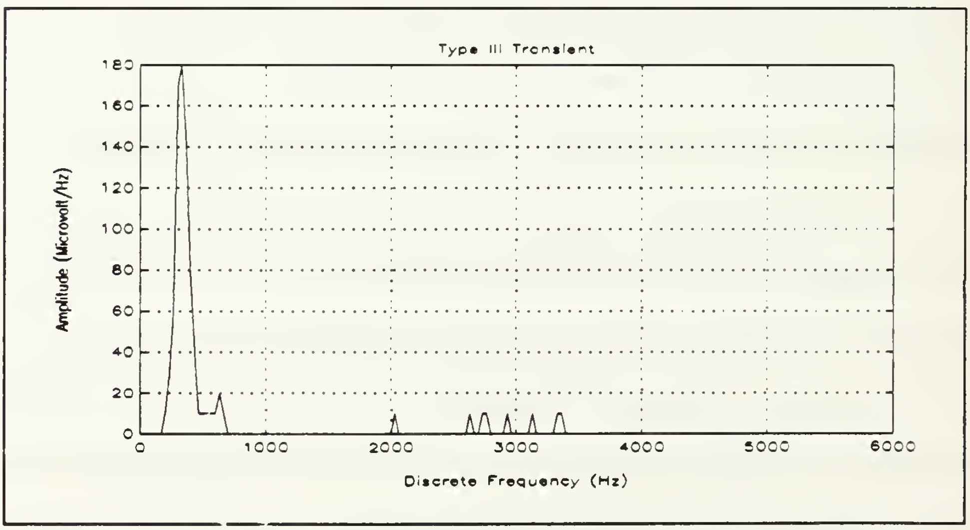

Figure 22: Type III Transient; Frequency Domain

Had the amplitude scale of Figure 21 been similar to those used in the other two figures the small peaks near 3000

Hz in Figure 21 may have been evident in the floor of the data as they are in the other two figures. The scale used in Figure 21 is driven by the amplitude of the maximum peak, which is significantly larger than the maximum peak for the other two types of signals.

Finally before discussing the performance of the frequency network which was built and tested consider Figure 23 below which is a plot of all type I transients. A comparison with its time domain counter part will show that although variability does exist there is significantly more structure here than in the time domain, owing to the frequency domain advantages previously discussed.

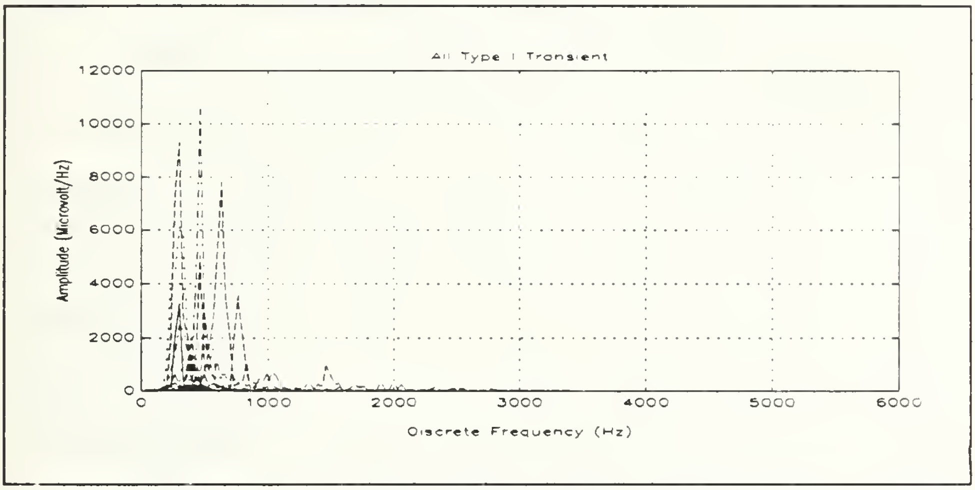

Figure 23: All Type I Transients; Frequency Domain

2. *Results: Testing the Frequency Domain Network*

For this portion of the testing a number of networks were built and tested. The basic network consisted of 150 neurons at the input layer, a hidden layer with 60 hidden neurons, a second hidden layer with 15 neurons, and an output

layer with 3 neurons. This network learned the training patterns to less than 0.01 rms error in 150,000 cycles of training. Training beyond 150,000 cycles failed to provide any further significant reduction in error so the network was tested. Test results were 60% recognition of type I signals, 50% recognition of type II signals and 25% recognition for type III signals.

As the performance of the basic frequency network was somewhat disappointing two additional enhancements were made to attempt to improve network performance. First a review of Figure 20 or Figure 23 shows that for the most part all of the signal information is contained in the first 1500 Hz of the record. As a first attempt at improvement, the long tails of comparatively little information were removed leaving a record spanning the range 0-1730 Hz. This reduced the size of the individual vectors from 150 to 52 points. A network with 52 input neurons, a single hidden layer with 25 neurons and an output layer with 3 neurons was trained for 60,000 cycles. Rms error again became slightly less than 0.01 and stabilized such that further training did not significantly reduce error. This network was then tested with recognition results as 73.3% for type I, 75% for type II ,and 25% for type III.

Last, the records were reduced in size in the time domain to 256 points and then split in half in the manner discussed in the data enhancement techniques to produce two records of 128 points each. These records were then

transformed to the frequency domain and the redundant second half of the signal discarded. This procedure had the effect of doubling the data while still retaining independence. Figure 24 represents a typical type I transient, the records produced from Figure 24 are provided below as Figures 25 and 26. Comparison of these figures reveals that although the two reproduced signals are somewhat different from the "parent" signal they are sufficiently like one another to allow the network to adequately train on both as type I signals. For example both show peaks at 400 and 700 Hz and valleys at 550 Hz albeit the magnitude is variable between the records.

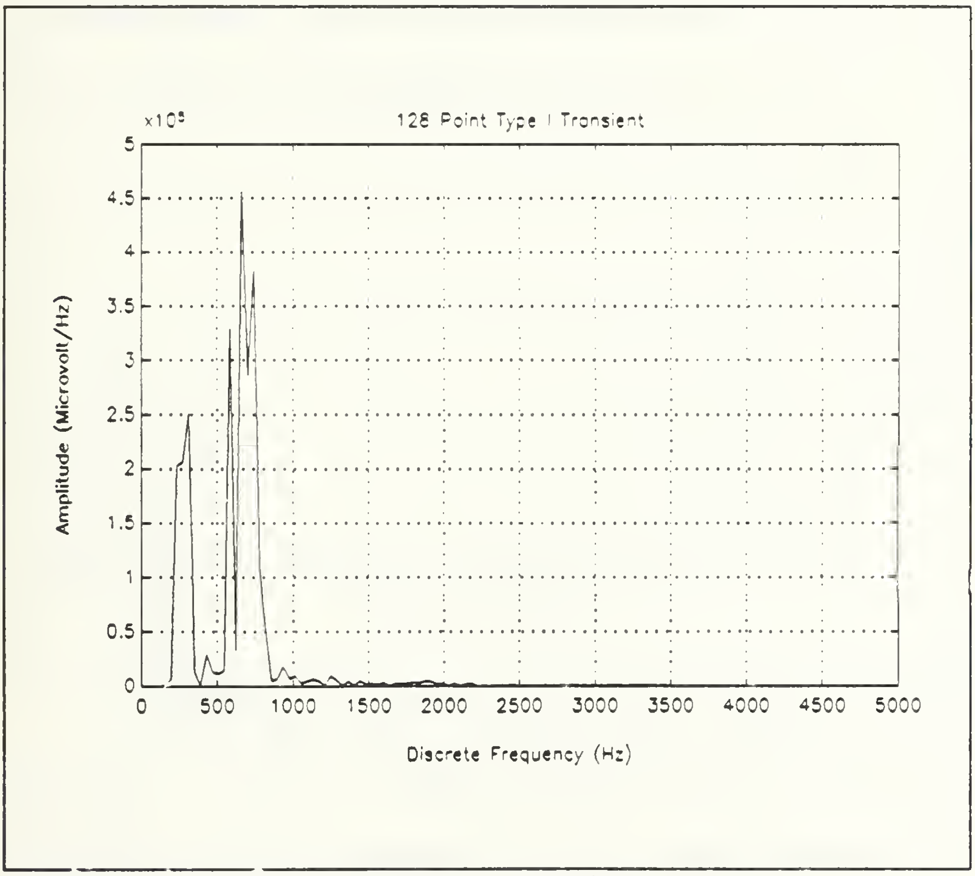

Figure 24: 128 Pt Type I Transient

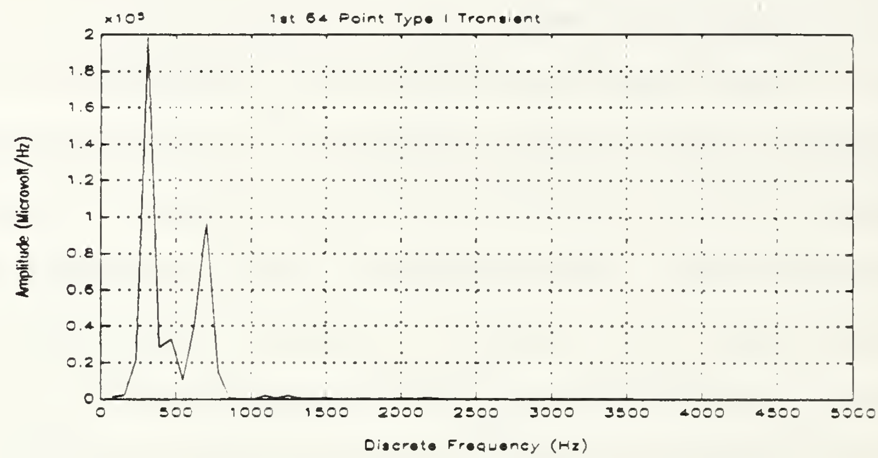

Figure 25: First Exemplar From Fig 24 Data

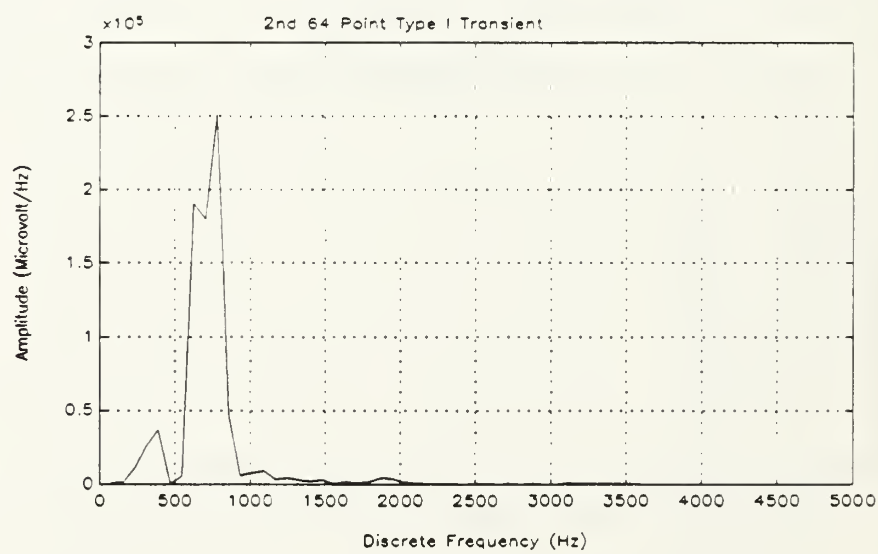

Figure 26: Second Exemplar from Fig 24 Data

A new network consisting of 64 input neurons, 20 neurons in the first hidden layer, and 12 neurons in the second hidden layer, with 3 output neurons was built and tested. This

network provided the best and most consistent results in the frequency domain. Performance was 83% recognition of type I transients, 75% recognition of type II transients, and 25% recognition of type III transients.

As can be seen all networks in the frequency domain performed poorly in recognizing type III transients. Type III transients are those transients associated with biologic noise in the ocean. Figure 27 shows the four type III transients used in the test data file which the networks were asked to classify. Only the first third of the signals has been graphed (0-1667 Hz) because the signal amplitude virtually disappears past approx 1500 Hz and this scale makes variability easier to discern.

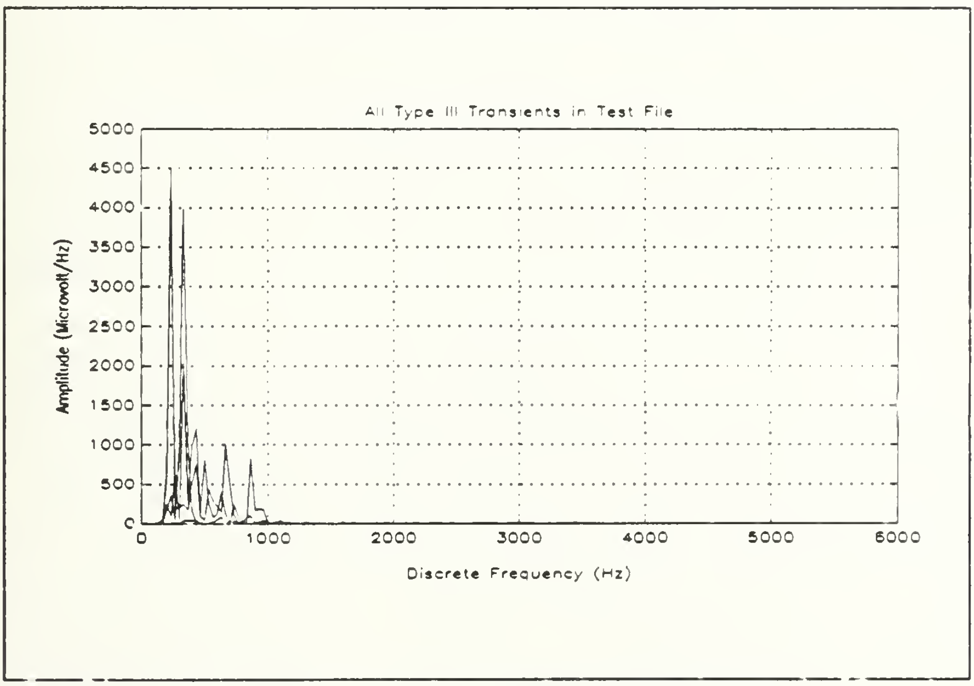

Figure 27: Test File Type III Transients

Nothing more is known about the original source of the biologic noise. Thus it is quite conceivable that the first record could be from a dolphin while records two, three, and four might be from entirely different sea mammals or fish. As previously explained neural networks are capable of making these types of generalizations but must have sufficient data to do so. In this case there simply exists too much variety in too few records for these networks to properly generalize. This it is believed accounts for the consistently poor performance of type III transients.

V. REDUCED SIZE FEATURE BASED CLASSIFIER

A review of the previous two sections would indicate that a feature based neural network classifier is feasible. In fact given the complexity of the acoustic transients to be classified it would appear that this type of classifier is preferable to one which classifies in the time domain or frequency domain. Clearly the performance of the network which classified on 31 independent features was superior to those classifying in the time of frequency domains. For example, for type I data, Table 2 shows that the feature based network recognized 92% of type I transients while the time and frequency domain networks of section IV only recognized 60% and 83% of type I transients respectively. This comparison leads one to consideration of again utilizing a feature based network but reducing the size of the network. Investigations into reducing the size of the feature based network are now considered.

A. ADVANTAGES OF A REDUCED NETWORK

One advantage of a reduced network is the increased speed with which a network can respond. The significance of this analysis and the subsequent reduction in network size it produces is immediately apparent from review of Figure 28. Figure 28 is a graph of the number of multiplications per training cycle necessary to update a three layer network which

is fully interconnected and learning via backpropagation as a function of network input layer size.

This figure is based on a single hidden layer that is 80% the size of the input layer.

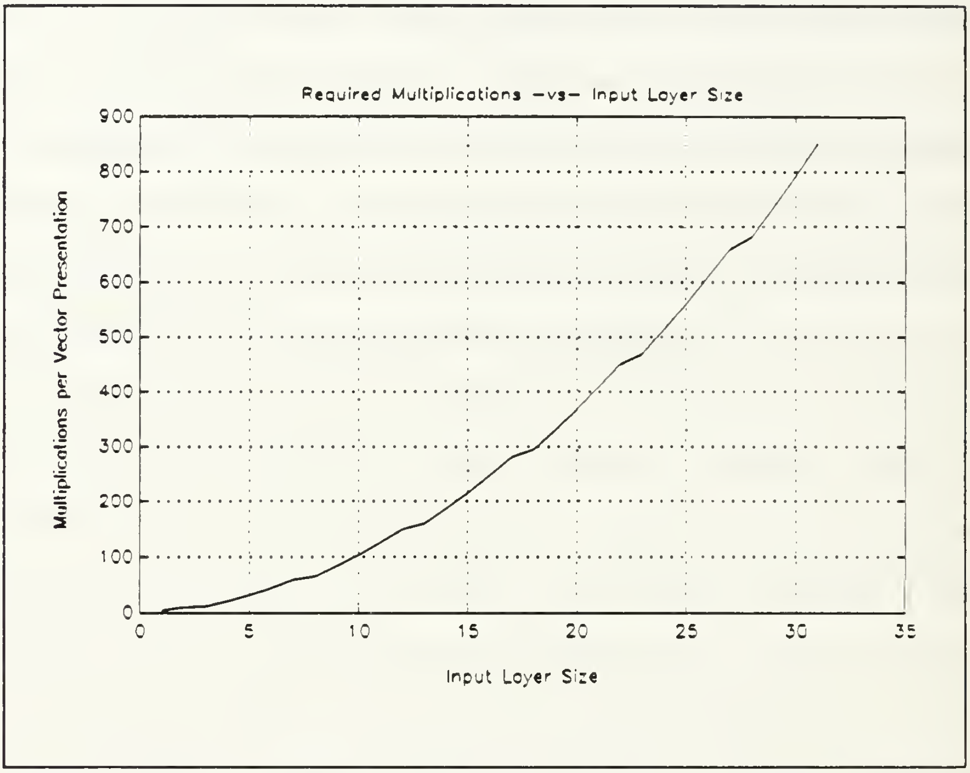

Figure 28: Training Time -vs- Network Size

These values correspond well to the final network presented in section III, which was 31 input neurons, 25 hidden neurons in a single layer, and 3 output neurons. As can be seen from Figure 28 this network would require 850 multiplications per input vector to conduct weight update. However a reduction in the input layer of only 10 neurons (total now of 21) results in only 400 multiplications. Thus for a 33% reduction in network input size training time is

more than halved. In any real world application it is not the training time which is of primary consideration but rather processing time during recognition. The networks discussed in this section do produce reductions in recognition time, although recognition times for both the full size feature based networks of section III and the ones considered in this section are on the order of microseconds.

Most significantly, another advantage of making these investigations in reducing network size, is that it allows one to determine which features are actually being used by the network to make the classifications and distinctions between different data types. This can be important because it reduces the amount of data which must be collected and later processed, yet still provides for reliable recognition.

B. FEATURE ANALYSIS

As a means of addressing the question above it is necessary to look at the individual records in detail and try to discern which parameters or features in the records characterize the information in the signal. There are fundamentally two approaches to this type of analysis. The first type of approach is theoretical in nature, and seeks to strongly establish underlying unique features of the signal. Several researchers have conducted these types of investigations. One particularly good investigation of this type is found in the Journal of Underwater Acoustics [Ref .9]. The second type of investigation is empirical in nature. The

analysis which proceeds here is of the second type.

One clue that the signals might contain redundant information is the singular value decomposition that was done in section III. Recall that this analysis led to the conclusion that there were approximately 21 independent variables in the combined data sets. See Figure 9. Thus it might seem reasonable, as a start to identify the ten input features which are not independent and eliminate them.

The software used to produce the neural networks in this thesis is a commercial product distributed by Neural Works Inc, entitled "Neural Ware Professional II Plus". One feature of this software is the ability to examine individual weights to and from individual neurons during and after training. Thus as a first attempt at reducing network size, the 31 x 25 x 3 network described in section III was trained for 220,000 cycles and individual weights were examined. In particular input connections to the hidden layer, which contributed less than 1% of the mean input, were searched for as possible candidates for deletion.

The search of the 31 x 25 x 3 network provided 13 candidates for deletion, these being feature number 2, 11, and 16-26. These features were first explored by removing these inputs and retesting the original 31 feature test set. This testing did indeed reveal that the deleted features were contributing very little to the overall recognition of the vectors in the test set. This was encouraging but it should be

noted that this network was still trained utilizing all 31 features, thus any potential savings in training time were not realized as discussed above.

Next the candidate features were actually deleted from the training and test files. This resulted in training and test files which were 18 column vice 31 column matrices. A new network was built which contained 18 input neurons, 15 hidden neurons in a single layer, and three output neurons. This network was trained for optimum recognition, 220,000 cycles, and tested. Results were 88% recognition for type I vectors, 95% recognition for type II vectors, and 96% recognition for type III vectors.

This network performance compares well to the recognition percentages given in section III. Type II and III data recognition is roughly equal for the two networks and Type I data only experienced a 4% reduction in recognition (0.88 down from the 0.92 for the full size section III).

Given the success of this process, the 18 x 15 x 3 network was examined for analogous reductions and 3 additional candidates for deletion were identified. These features were # 3, #12 and, #27 of the original 31 features. Deletion of these features led to a 15 x 12 x 3 network. This network was tested and led to the following recognition percentages: 88% for type I, 55% for type II and 95% for type III. Further attempted reduction in the size of the network resulted in serious degradation in performance.

Comparison of the above data suggests that this type of task can be reliably performed by a 18 x 15 x 3 network. This network trains and recognizes in less than half of time of the original feature based network yet still maintains an average recognition percentage which is above 88% for all data types.

One additional consideration with this network is the reduced signal pre-processing time. Details of the signal processing necessary to extract the relevant features has not been provided here. Suffice it to say that some of the features do require significant signal processing to extract. The benefits of reducing the number of features extracted from the original 45 provided by NSWC to the final 18 utilized in this successful network is obvious.

Further this analysis demonstrates that indeed the information content of a random extremely short duration transient can in fact be described in just a few data parameters. Undoubtedly, which features contain the majority of the information is directly related to the nature of the transient itself.

Again then a practical use for a neural network is demonstrated in the field of acoustic processing. This question of signal parameterization and classification or sub-classification is a very complicated one. The neural network demonstrated here rapidly extracted the information, by separating the features into those which actually characterize the signal and those which were redundant or did not contain

much signal information. This would be important information for those involved in actual data collection to have apriori, because it greatly simplifies the data collection task.

C. RESULTS: TESTING THE REDUCED NETWORKS

Table 4 below summarizes the pertinent information contained in this section by providing a side by side comparison of the two networks considered here with the final network considered in section III. The networks listed in each row of Table 4 are indexed by the following list of size and network dimensions.

1) Network #1 = 18 x 15 x 3
2) Network #2 = 15 x 12 x 3
3) Network #3 = Section III network: 31 x 25 x 3

The Table 4 column labeled "Normalized Training time" is given in arbitrary units and represents the number of floating point operations necessary for the computer to carry out its instructions in updating the weight matrix, normalized to one for the largest network. Thus if it takes 10 minutes to train network # 3 on machine "x" then it will take 3.7 minutes to train network #1 on the same machine.

In reviewing Table 4 note that smaller (18x15x3) network #1 (row one) achieved recognition percentages (0.88,0.95,0.96) which were nearly as good as the recognition percentages (0.92,0.94,0.95) for the much larger (31x25x3) network in row three. This might seem puzzling in light of the singular value analysis done in section III. A closer look indicates the

number of misclassifications actually did go up with net #1 when compared to net #3. A review of Table 1 shows that the 199 test vectors were distributed as 86 type I, 33 type II, and 80 type III. Thus the percentages in row one above represent a total of 15 misclassifications while the percentages in row three represent a total of 13 misclassifications.

TABLE 4: REDUCED NETWORK TESTING RESULTS

Network Comparison	Type I Recog %	Type II Recog %	Type III Recog %	Normalized Training Time
Net # 1 (18x15x3)	.88 (76/86)	.95 (31/33)	.96 (77/80)	.37
Net # 2 (15x12x3)	.88 (76/86)	.55 (18/33)	.95 (76/80)	.25
Net # 3 (31x25x3)	.92 (79/86)	.95 (31/33)	.95 (76/80)	1.0

Nonetheless the data suggests that yielding just a few additional misclassifications can result in a significant reduction in overall network size and training time. More importantly, if a 4% reduction in recognition percentage is acceptable for the particular application, significant reductions in data collection can be realized. Additionally,

much of the required (and very time consuming) data pre-processing asociated with the feature extractions can be avoided.

VI. CASE STUDY: THE NEURAL ACOUSTIC INTERCEPT RECEIVER

A. BACKGROUND

Up to this point the type of signal considered in this thesis has been a random unintentional short transient, i.e. transients on the order of 10 msec or less. As a final consideration it is desirable to look at the neural network as an active intercept receiver.

The need to intercept and classify underwater active sonar is well established. Needs vary from biological applications such as fish population counting to military applications such as active sonar analyzers for submarines. As a submarine relies on stealth to fulfill its mission, the acoustic intercept receiver when properly employed is indispensable to maintaining this stealth. Like many warning devices it must be capable of providing warning sufficiently in advance to allow the host submarine to maneuver and thus avoid being detected by acoustic means.

B. PROBLEM SETUP

The problem considered here is fundamentally a different one than the problem traditionally considered by transient detection researchers, namely that of extracting and classifying short unintentional transients. This fact arises from the differing nature of the signal. Unintentional signals are generally extremely short in duration and somewhat random

in nature both in the time domain and frequency domain. Additionally signal to noise ratios are quite small. All of these contribute significantly to the difficulty of the classification task and the need to conduct feature extraction and signal processing to get reliable classification results. The nature of the intentional active sonar transient is considerably different. Consider that the active signal Source Level for typical transmissions exceeds 220 dB re 1 μPa @ 1m, the signal is mono-frequency and stable in content, or at least is swept in a predictable pattern, and finally the signal duration is almost always in excess of 50 msec and often approaches 500 msec or more.

It should be apparent that these features are exactly the ones which make the detection of short unintentional transients so difficult.

To examine this problem two different cases were considered. First an application is considered which would consist of the network being utilized as a stand alone intercept system which receives input from the FFT of the broadband times series energy and is expected to classify signal frequency content and other appropriate signal parameters. In the second case a neural network is considered as an adjunct classifier to a traditional acoustic intercept receiver. In this case the network is expected to use the intercept receiver signal parameters as input and make specific sonar type classifications.

C. THE STAND ALONE NEURAL INTERCEPT RECEIVER

1. Background Physics

To study this problem effectively it is necessary to define the parameters with which such a system must operate. Characterization of these parameters will allow training and test data to be built that can assess in a fair manner the performance of the neural network acoustic intercept receiver when compared to traditional systems.

It is assumed that the system must be capable of providing reliable recognition and classification at a range which would provide a very low acoustic probability of counterdetection for two platforms operating within the same homogeneous ocean. The passive sonar equation in its simplest form is:

$$SL - TL = NL - DI + DT \tag{16}$$

Active sonar detection includes two cases [Ref. 2]. In the first case the environment is considered to be reverberation limited and in the second the environment is noise limited. The only case considered here is the noise limited environment. In the case of the noise limited environment the active sonar equation can be written as:

$$SL-2TL+TS \geq NL-DI+DT \tag{17}$$

Where

SL = Source level of the active sonar

TL = The transmission loss between the source and target

TS = The nominal target strength of the target

NL = The noise present in the spectrum considered

DI = The directivity index of the processing system. This really represents the systems ability to gain performance by discriminating against the noise field in a given direction

DT = The detection threshold. This represents the amount of signal excess required for an operator to make the decision that a valid return is present

Analytical definitions of each of the above terms are widely available and the standard definitions are used here [Ref. 2][Ref. 8]. However the Detection Threshold plays such a key role in this type of detector that further elaboration is provided.

The Detection Threshold is a performance measure of the system, defined as :

$$DT = 10 \cdot \log \frac{S}{N} \tag{18}$$

Where

S= Signal power

N= Noise Power

but can also be expressed in terms of the detectability index

"d′" the system bandwidth "w" and the pulse duration "τ" as:

$$DT = 5 \cdot \log\left(\frac{(d')^2}{w \cdot \tau}\right) \tag{19}$$

In this form "d′" is the detectability index, which is related to the classic detection index "d" through $d=(d')^2$ [Ref. 8].

When establishing problems of this nature there always exists a tradeoff between probability of detection and probability of false alarm. In an environment rich in active sonar, biologics or other types of transient noise the false alarm rate must be controlled. The criterion adopted here for these competing interests is that the active emission must be classified 95% of the time at a range equal to or exceeding the range corresponding to 5% probability of counterdetection, while not exceeding 5×10^{-2} false alarm probability. This formulation gives rise to a set of receiver operating curves of the form given below in Figure 29 [Ref. 8]. These receiver operating curves represent the operating charachteristics for a detection system whose probability of detection and false alarm probability are distributed as Gaussian with equal standard deviations.

Review of Figure 29 shows that the system described, given the constraints on probability of detection and false alarm rate, is required to operate at a detectability index of four, marked on Figure 29 as the "Operating Point".

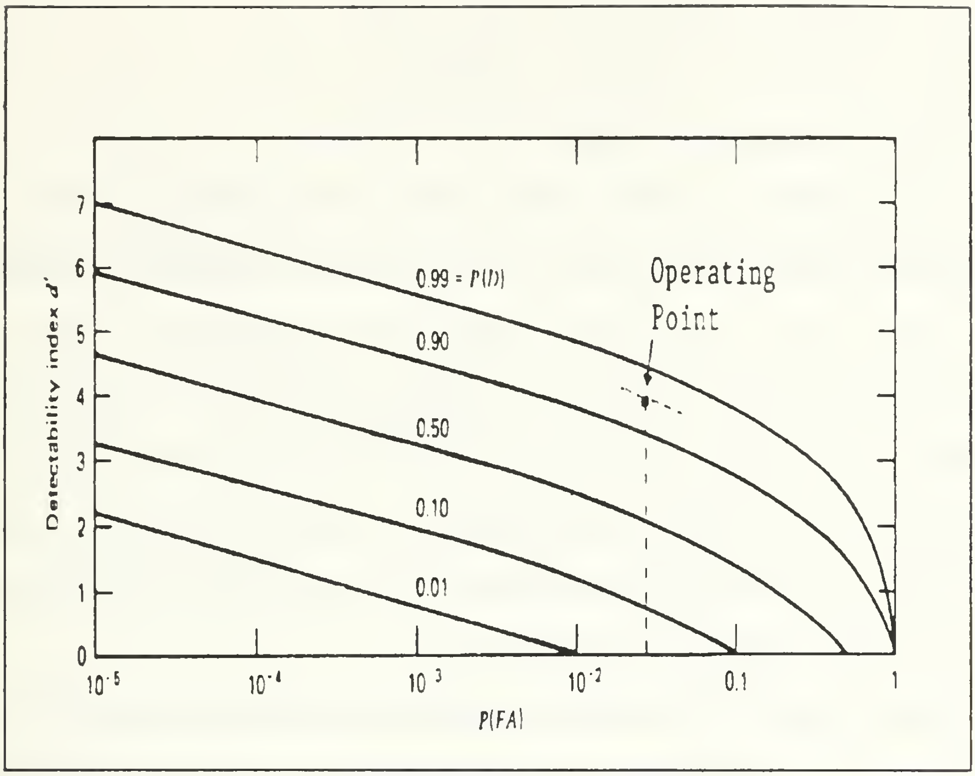

Figure 29: Detectability Index Curves

2. Data Formation

The data set established for the first case consisted of four different types of data. This data consisted of a low frequency threat signal, a band of low frequency detections which are not considered threat, and analogous high frequency signals. This data breakdown is consistent with that processed and displayed by traditional acoustic intercept receivers.

The "threat bands" consist of detections at a single frequency while the non-threat "detection bands" cover a wide range of frequencies and would be activated for any detection in the band. The frequencies picked for this study are:

1: Low Frequency threat: 1.1 kHz

2. Low Frequency detect: 1.5-1.9 kHz

3. High Frequency threat: 3.6 kHz

4. High Frequency detect: 3.3-3.8 kHz

Note that the low frequency threat lies outside the low frequency detection band but the that the high frequency threat lies in the high frequency detection band. The implications of the latter formulation are that if a signal in the band 3.3-3.8 kHz other than 3.6 kHz is presented to the network an "HF DETECT" output should be processed but if a signal of 3.6 kHz is presented to the network then an "HF THREAT" output should be processed.

One important question which must be addressed is the amplitude of the frequency components relative to the noise field to make the problem characteristic of actual conditions yet still meet the detection index and threshold requirements. This question is answered by evaluating the underlying physics of the sonar equations and the constraints of the problem.

Assuming a noise limited environment, solution of Equation 17 for Source Level yields

$$SL = 2TL - TS + NL - DI + DT \tag{20}$$

For a homogeneous layered ocean with both source and receiver in the sonic layer a simple model for transmission loss becomes:

$$TL = 10 \cdot \log(r) + a \cdot r \tag{21}$$

The absorption coefficient "a" in Equation 21 is strongly a function of frequency, and can be approximated by:

$$a = \left(\frac{8 \times 10^{-5}}{0.7+f^2} + \frac{.04}{6000+f^2} + 4 \times 10^{-7}\right) f^2 \quad \frac{Db}{m} \tag{22}$$

over the frequency range of interest here for most high power long range active sonars, provided frequency is in kHz [Ref. 8].

The detection probability in Equation 20 is hidden in the detection threshold term. We are interested in the Source Level at which a 5% probability of detection occurs. This Source Level of course depends on all of the terms of the equation, but if all terms are kept constant at nominal realistic values such as those proposed by Urick it is possible to determine the Source Level (noise limited environment only) of the tone required to make this detection [Ref. 2]. Interpolation of Figure 29 shows that for a detection probability of 0.05, and a false alarm probability of 10^{-3}, the required detectability index is approximately two. Given a signal processing time of 500 msec (reasonable for an active sonar receiver) and a bandwidth of 100 Hz centered at 1000 Hz (reasonable for doppler associated with modern day submarines) Equation 5.3 yields a Signal to Noise ratio of

0.28 or -5.5 dB.

Figure 30 presents mean values of the deep ocean ambient noise spectrum level for 10-20,000 Hz [Ref. 8].

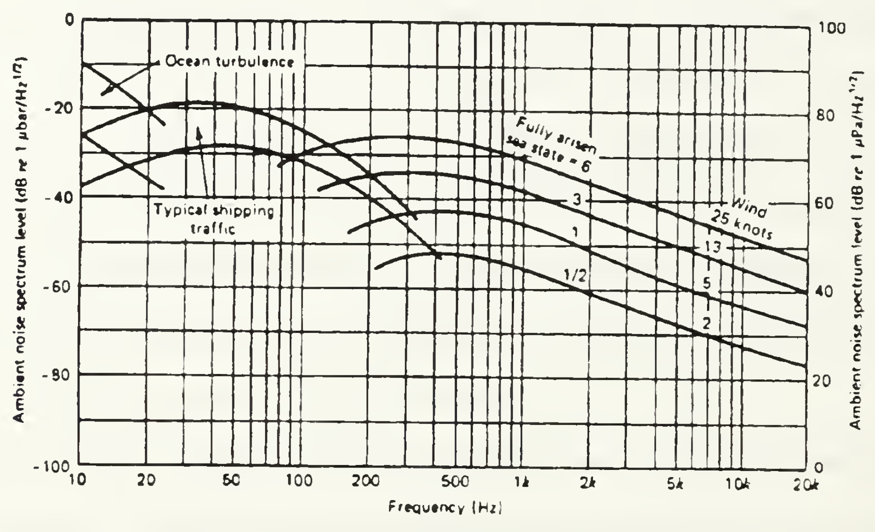

Figure 30: "Wentz" Ambient Sea Noise Curves

It is seen that ambient noise near 1000 Hz is approx 62 dB for a sea state 3. For purposes of this discussion it will be assumed to be 62 ± 3 dB re 1 μPa in the 100 Hz band around 1000 Hz. This being the case and assuming a nominal range of 20,000 m Equation 20 yields a source level of 207 db re 1 μPa @ 1m to make this detection.

To obtain the final signal power in the frequency bin of interest this source level is attenuated through 20,000 m of range (one way trip), and then processed through a 100 Hz filter operating at 1000 Hz from a square law detector. Next the total band noise level with the tone absent is calculated from:

$$BL = PSL + 10 \cdot \log(w) \tag{23}$$

The Pressure Spectrum Level (PSL) in Equation 23 is simply the ambient noise field near 100 Hz and is again assumed to be 62 dB. Finally the SPL of the tonal is logarithmically added to the noise spectrum to ascertain the final total band level. Omitting details of calculations this number turns out to be 106 dB. This number represents the level of the signal at the detecting platform and provides the basis for building the signal part of a data set to test neural network reliability, recognition, and classification as an acoustic intercept receiver under the stated detection and counterdetection constraints.

It is recognized that the required source level calculated here is highly dependent on range and the assumption that the environment remains noise limited. The noise limited assumption is rarely met throughout all ranges but is used here as a simplification necessary to solve a standardized problem. With respect to the range question, if the range were to double then a new required source level would result, this then could be attenuated as before through half of the range and a new sound pressure level of the tone at the target submarine would result. This process is highly non-linear, the nominal value of 20,000 m was chosen to provide a consistent basis for making comparative evaluations

of the neural network performance.

The foregoing discussion builds one data point, namely that centered in the 100 Hz band just above 1000 Hz. To form an entire data set one needs to repeat the process through the entire range of interest, reformulating the problem in terms of different ambient noise, and incorporating the frequency dependence of the other frequency dependent terms of Equation 20.

Data were built based on the physics described above. Figure 31 is a representative exemplar that would be provided to the network for recognition. This figure represents the energy resident in each of 30 frequency bins. This energy is found by integrating all of the noise intensity over the width of the band and then displaying the entire band as the average value of the integration.

Note that Figure 31 contains a signal at 1100 Hz and also that the noise is not constant with frequency as reflected in Figure 30. This particular exemplar is the frequency used to simulate a low frequency threat sonar. Further note that Figure 31 consists of a total frequency range of 1000-4000 Hz. With a 100 Hz bandwidth this corresponds to 30 separate input bins, and thus sets the size of the input layer of the neural network at 30 neurons. The data are presented here, in the energy spectrum formulation mentioned above, as they would appear after band level processing.

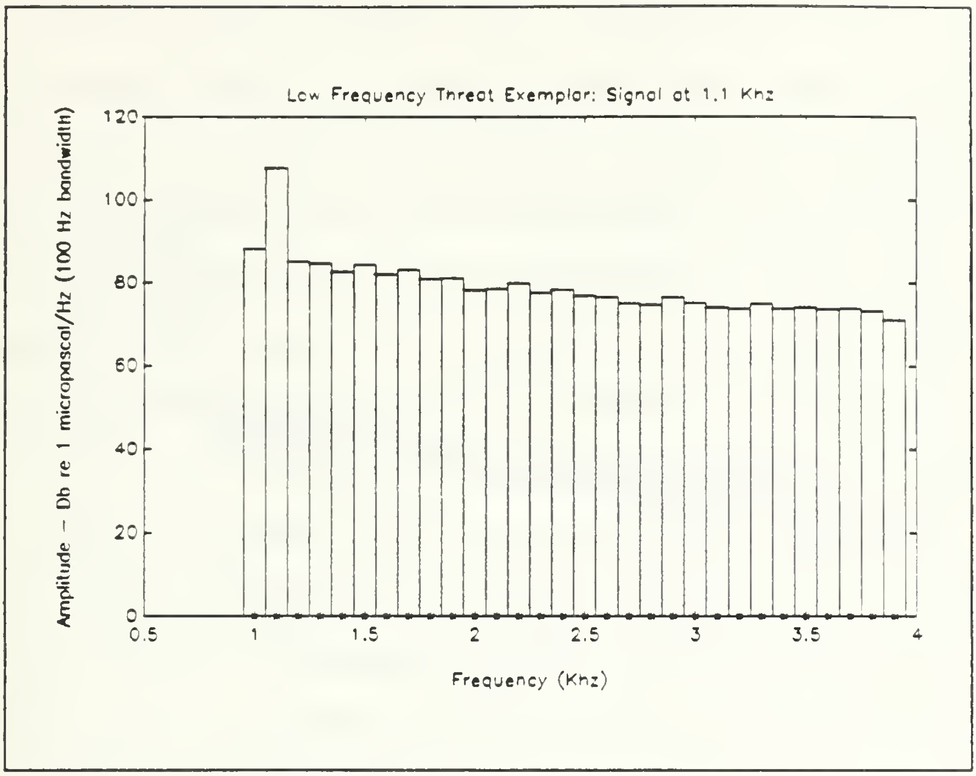

Figure 31: LF Threat Exemplar; Band Level Processed

Review of the physics which led to the choices in bandwidth and frequency coverage here point to important tradeoffs when building a network expected to function over a large frequency range. From equation 23, as bandwidth becomes smaller the total band level also goes down, and more importantly the contribution of the tonal to the energy in the band becomes proportionately larger. Thus smaller bandwidth would seem better, however if bandwidth was reduced to 10 Hz, for example, then coverage of the same frequency range requires an input layer size of 250 input neurons. Thus the tradeoff is between a large network with smaller bandwidth and higher signal to noise ratios, and smaller network size which

requires fewer multiplications, but in turn means wider bandwidths, and thus lower signal to noise ratios (with corresponding decreased reliability in detection). Last, it should be noted that the average noise field appears in Figure 31 at approximately 20 dB above the 62 dB previously derived. This additional 20 dB arises from the band level processing which results in the integration of the noise field over the bandwidth, i.e. the 10 log(w) term in Equation 23

Multiple exemplars of each type of data were constructed utilizing the guidelines discussed above and the modifications explained below. Figure 32 shows the 50 exemplars of the low frequency threat portion of the training set.

Each exemplar was constructed from a "fundamental" exemplar with a small random spread about the fundamental for the data type. Note that the individual exemplars range from 1.05 to 1.15 Hz (because of the 100 Hz bandwidth) at 106 dB and signal amplitude varies from 103 to 109 dB. Amplitude variation was produced by adding a normally distributed random variation to the 106 dB signal and was picked to simulate real world variability in source level. This accounts for the fact that real sources do not produce exactly the same source level on every transmission. The construction of noise field data involved making an empirical fit to Figure 30 in the range 1 - 20 kHz.

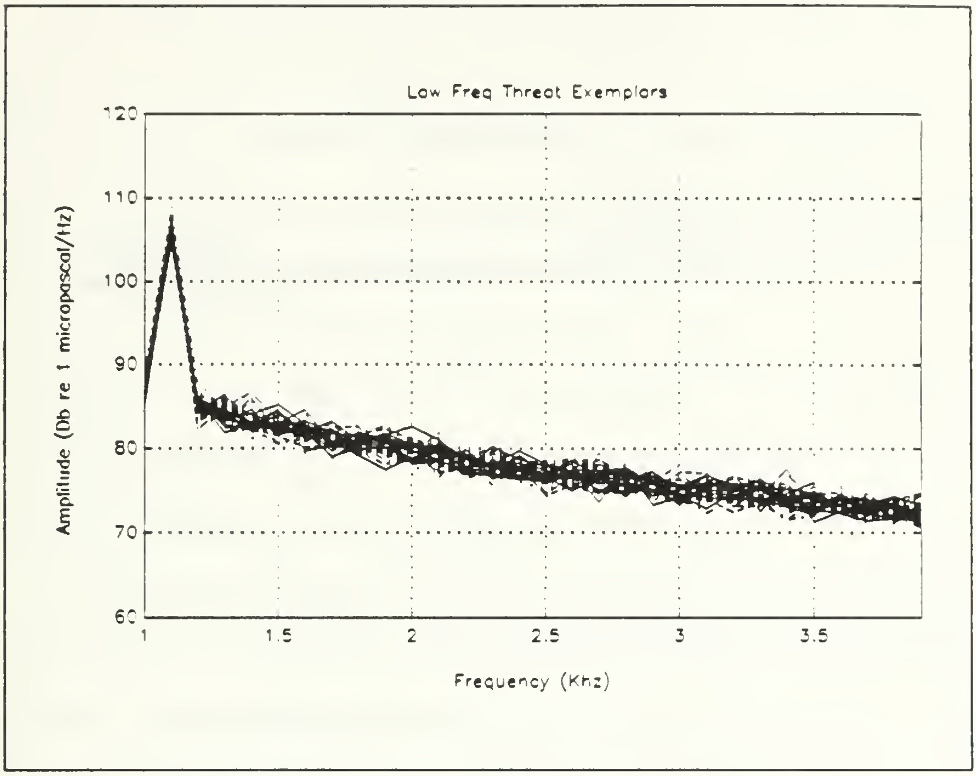

Figure 32: All LF Threat Exemplars

Review of Figure 30 shows the data to be plotted in a semilog fashion, implying an exponential relationship between noise in dB and frequency. This data was empirically fit to within 3% rms error by:

$$\textit{Noise Level} = A - B\cdot\ln(f) \qquad (24)$$

With A=67 dB, B=10.6 dB, and f in Hz.

Random variations of up to 3 dB, to account for sea state variations, were then added to the noise data generated by this empirical equation to yield the final noise data set. Four different signal types comprise the data set. The entire

training data set is presented in Figure 33.

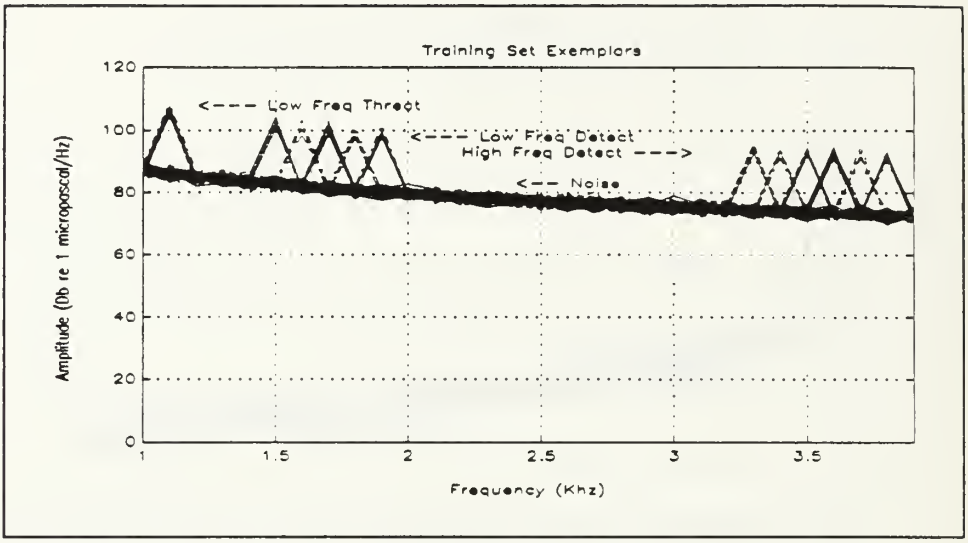

Figure 33: Entire Training Data Set

The high frequency threat data is not explicitly labeled on Figure 33 as it is contained within the high frequency detect band.

3. Results: Testing the Stand Alone System

A backpropagation network incorporating generalized delta rule learning was constructed and tested with the data prepared as described. The goal of the testing was to ascertain the ability of a feed forward neural network in recognizing mono-frequency signals of sufficiently low amplitude that the output could be used reliably as an early warning acoustic intercept receiver. A secondary goal consisted of examining the ability of the network to determine some representation of the amplitude of the signal being presented.

Data built and described above were split in half to

form independent training and test sets. These data were presented to a neural network consisting of a 30 neuron input layer, a 15 neuron hidden layer, and a 4 neuron output layer. The network was trained to an rms error of 0.01. The network was then tested with the following results:

1) Low frequency threat recognition: 99%
2) Low frequency band detection recognition: 96%
3) High frequency band detection recognition: 96%
4) High frequency threat recognition: 100%

This data suggests that a neural network can reliably (> 95%) recognize signals which are resident in a noise field with signal to noise ratios comparable to those which would result in 5% counterdetection probability.

False alarm probability was assessed by constructing a separate data test set which contained 1000 exemplars of noise only. The network was trained on the original training set (which contained no exemplars of noise only) and then tested on the "noise" data set. A false alarm was judged to have occurred if any output neuron exceeded 0.8 activity level. False alarm rate by this method was 5×10^{-2}.

To achieve these detection and false alarm rates the system output neuron activity to provide a valid detection was set at 0.89. This value provides the optimum tradeoff between high detection rates, which go down as this value is increased, and false alarm rate, which also decreases as this value is increased. Review of Figure 29 shows this system to

be operating at the desired detectability index of four.

The secondary goal of this research was to assess the networks ability to further parameterize this data, ultimately for output display. The single most important feature which needs to be assessed is the strength of the incoming signals. Signal strength forms a basis for assessing counterdetection vulnerability.

Figure 34 is a graph of the signal portion only of the 100 LF THREAT signals resident in the test set that was presented to this neural network.

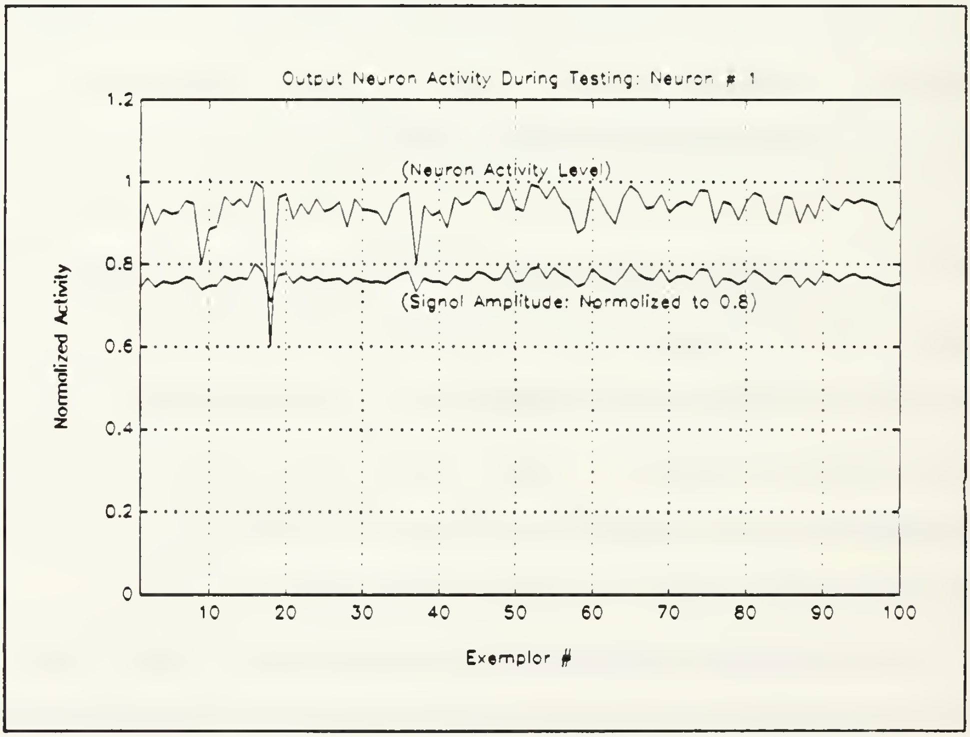

Figure 34: Neuron One Activity during Testing

Graphed with these signals is the corresponding output activity level of output neuron #1 as the input vector was being presented to it. A typical value for the input signal

level would be 106 (dB re 1 μPa) but these values have been normalized to a maximum value of 0.8 so that they may be displayed on the same graph.

Output neuron activity is already normalized. Figure 34 suggests that input signal level and output neuron level are highly correlated. Correlation coefficient from this data when regressed linearly was 0.88. Thus it appears that signal strength determinations are in fact achievable from information resident in the neural network.

Other signal parameters which may be of interest include signal relative bearing, period between pulses, and signal duration. Relative bearing of the signal is a function of the directivity of the sonar hydrophone not the signal processing and as such is not considered here. Signal duration and period between pulses (sometimes known as threat period) can easily be obtained by utilizing simple counters at the input and output of the neural network but are not optimum tasks for the network itself to perform.

D. THE ADJUNCT INTERCEPT RECEIVER

As an alternative approach to stand alone acoustic intercept this research also considered a simple neural network as a supplement to a traditional acoustic intercept receiver. In this case the network is presented with a small set of features which have already been extracted by a traditional intercept receiver and is expected to provide classification of the signal.

This sort of problem is fundamentally different from the previously considered problem because in essence the inputs to the network form a very small set (3 in the work conducted here) and the possible outputs may be quite varied and large in number. This type of problem has been extensively studied by McClelland and Rumelhart with respect to interactive activation and competition [Ref. 4]. The approach considered here is again to apply the backpropagation methods utilizing supervised learning to this classification task.

1. Data Construction

Data for this examination contained the following three inputs: Signal frequency, pulse length, and threat period. Table 5 below summarizes the base values for these different signal types. All parameters are fictitious.

TABLE 5: FEATURE BASED DATA

Feature based Data Summary	Frequency (kHz)	Pulse Length (msec)	Threat period (sec)
Submarine	2.5	300	5
Surf Warship	7.0	500	10
Torpedo	30	300	2
Sonobuoy	10	200	120
Biologic #1	17.0	500	Random
Biologic #2	45.0	10	Random

Data were constructed for six possible sources: submarine sonar, surface warship sonar, torpedo homing sonar, active sonobuoy sonar, and two distinctly different types of biologic noise. In addition to the basic data, each data type was constructed with two variants. For example in the submarine sonar case pulse length was changed to 250 msec for one variant and threat period was changed to 60 sec for the other. These variations complicate the classification task by requiring the network to classify all submarine transmissions as "submarine" regardless of which variant is presented. Also note that the threat period column of the biologic noise is listed as random. This field was obtained by generating random numbers corresponding to the range 1-1000 sec, as might be expected from biologic noise. Five exemplars of each variant was included in the training and test sets for a total size of 90 x 3 for each set.

2. Results: Testing the Adjunct System

A 3 x 3 x 6 neuron backpropagation network was built utilizing generalized delta rule learning. The network was trained to minimize rms error and tested. Results are reported in Table 6. Table 6 recognition results are provided for two different detection criteria. In method A output neuron activity of 0.8 or greater results in reporting a valid detection. Method B results are reported as correct if output neuron activity for the associated sonar type exceeds that for the other output neurons.

TABLE 6: FEATURE BASED NETWORK RESULTS

RECOGNITION PERCENTAGES	Method A Criterion	Method B Criterion
Submarine	100 %	100%
Surf Warship	67 %	100 %
Torpedo	100 %	100 %
Sonobuoy	33 %	73 %
Biologics #1	0 %	100 %
Biologics #2	100 %	100 %

When interpreting Table 6 results recall that the test data set was small. Detection results represent the percentage of successful detections made in 15 opportunities. Using method A detection criterion to grade false alarms resulted in a false alarm rate for the entire data set of zero. A false alarm is again considered to have occurred when an activity of 0.8 or greater results for an output neuron other than the one intended for the signal being tested.

VII. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

A. SUMMARY

The goal here has been to present neural networks as a new and promising approach to transient classification. Their power lies in the ability of the network to generalize and to use features as a basis for optimum decision making in signal classification. This work holds great promise for application aboard U.S. Navy Submarines where this technology could be adapted to provide audible output of the decision making process and thus free up watchstanders who are now making these types of simple decisions.

This thesis has presented a neural network approach to the classification of active transmissions both intentional and unintentional. This type of classification is exemplary of the type which is necessary for a submarine to fulfill its mission whether it be transient signal processing or active acoustic intercept as an early warning detection device. Several systems have been explored.

First a backpropagation network was considered as a feature based classifier of unintentional transients of short duration. This was then compared to analogous transient processing in the time and frequency domains. Following this comparison a reduced size feature based detector was demonstrated which performed to within a few recognition

percentage points of the full sized feature based detector.

Next, neural network technology was applied to the active intercept problem in a case study. In the first part of the case study the neural system was considered as a stand alone acoustic intercept receiver. In this formulation the network was given a large number of inputs relative to the expected number of output classifications. The network presented here was highly successful in performing this task over a limited frequency range. As a second consideration a backpropagation network was considered as an adjunct classifier to an existing traditional acoustic intercept receiver. In this case the network was given a small number of inputs and expected to classify the sonar by type, with the number of expected classifications in the library of possible outcomes becoming potentially quite large. This latter task is the process that a human operator would undergo to make the same type of classification. This last method has a particularly useful application aboard U.S. Navy submarines where often the watchstander most in need of the information cannot process the information visually because he is using his eyes to man a periscope.

B. CONCLUSIONS

Based on the research presented in this thesis it is concluded that neural networks can reliably perform the task of sonar transient classification. Additionally one can conclude from the data presented here that this task is

optimized when the data set has been parameterized into features which characterize the data set.

The highlight of this thesis was a 31x25x3 neuron feature based multi-layer feed forward neural network. This network was highly successful in recognizing acoustic transients which had been parameterized into features which served to characterize the structure of the transient. With recognition percentages exceeding 92%, it can be stated that this network can reliably perform a task which would be virtually impossible by a human operator, and it can perform this task in much less time than that required by traditional signal processing.

Given that feature extraction and presentation to a neural network results in reliable transient recognition, one searches for the fewest and best features to present. It should be clear that this decision is highly data dependent, nonetheless the singular value decomposition presented here provides an excellent analysis tool for addressing this issue. The singular value decomposition performed on the data set in this thesis suggests that at least 10 (30%) of the features could be ignored. The result of this analysis was a smaller network and reduced training and testing times. Another tool which can be utilized if available is a review of the weights being processed to and from individual neurons. This analysis led to the identification of a total of 13 input features which were eventually removed. The analysis above produced a

reduced size network which trained in less than half the time of the full sized feature based network. Although performance was slightly degraded for this network (15/199 misclassifications compared to 13/199 for the full size network). The reduced size of the network provides a tradeoff worth considering if small performance compromises are not germane to the intended application.

One final significant conclusion of the transient recognition research presented here is that to reliably perform generalizations in pattern recognition, a neural network works best from a large data set. In the case of the time domain network presented in section IV of this thesis the data set was simply too small for the network to reliably conduct pattern recognition. This small data set resulted in recognition percentages of less than 60% as compared to the feature based networks which performed at better than 88% recognition for all data types. One should not conclude from this study that the time domain holds no promise for further research in this area, but rather that future work will require a larger data set. Minimum data set considerations are discussed in the recommendations section below.

Finally, a specific case study of these concepts as they apply to the active acoustic intercept problem demonstrated that a neural network can be used on small data sets to reliably extract active sonar transmissions from a noise field. Within the limited constraints of the problem here, a

neural network can be used to make classifications of already intercepted and processed active sonar signals.

The highlight of this portion of the research was the stand alone neural acoustic intercept receiver. This system produced recognition percentages exceeding 95% for all four data types and achieved a false alarm rate of 5%. Significantly, this system was able to provide information on the amplitude of the activating signal. This information is considered absolutely crucial to a system which is to provide reliable early warning. The network presented as an adjunct intercept receiver did experience some difficulty in making the proper generalizations. This is attributed to two factors. First the data set on which it was operating was relatively small (90 total exemplars, or 15 of each of six different classes of data). Second, neural networks are not particularly good at solving this type of problem, namely one where combinations of just a few inputs produce a relatively large number of outputs.

C. RECOMMENDATIONS

This is a limited study in many respects, the results however suggest that neural network classifiers should be able to provide a viable alternative to existing techniques for classifying intercepted unintentional transients and active sonar pulses. This thesis looks at a limited number of possible applications of this technology to the problem.

It is recommended that the data set be enlarged to include

a much larger feature based data set. This thesis looked at recognition of three different types of signals. The number of different data types should be expanded to all those which might be reasonably encountered in the real ocean environment. This will provide assurance that a feature based network can successfully operate over the wide range of input type data that might be expected in an actual shipboard application.

Additionally, one of the most significant limitations in the time and frequency domain was the limited availability of data. Accordingly it is recommended that this problem be re-studied with a significantly enlarged data base. One method of addressing the minimum size of a data set that might be appropriate, is to consider the sample size necessary to construct a 95% confidence interval from the results. This sample size is given by [Ref. 10]:

$$n=\frac{4\cdot(z_{\alpha/2})^{2}\cdot p(1-p)}{L^{2}} \tag{25}$$

Where

n = # of vectors in the data set

p = expected recognition probability

L = The length of the confidence interval

$z_{\alpha/2}$ = Value of the Standard Normal Random Variable

For the data described in this thesis we expect "p" to be near 0.9 and a reasonable value for L is 0.1. At 95% confidence $z_{\alpha/2}$=1.96. Putting these numbers into Equation 25

results in a data set size of 139 vectors. This number represents the number of vectors necessary to say with 95% confidence that a network is recognizing .9±.1 of the vectors in the set. This data set size does not in any way reflect the network's ability to perform recognition at this percentage, but rather to have confidence in the results if the network does perform to this recognition level. This data set size seems reasonable as a starting point in light of testing and conclusions presented for other neural networks in this thesis. Undoubtedly more data is always better, however given that unlimited data is not available this number provides a good starting point to achieve the type of performance standards expected in this type of recognition problem.

The data scales used in the acoustic intercept study have been completely arbitrary. The scales used could have been the 1-4 kHz, which was used, or could have just as easily represented 10-40 kHz. It is recommended that follow on work look at a greatly enlarged frequency range, for example 1-100 kHz. With a bandwidth of 100 hz this of course means a considerably larger neural network. Additionally the High and Low frequency detect regions should be enlarged to cover perhaps half of the band examined.

Further, it is recommended that follow on work include investigation of the active intercept problem in the time domain, as the time domain may provide the ability to extract more raw signal information from the neural network. For

example signal amplitude would appear to be reproducible again, utilizing output neuron activity level as a basis, such as the analysis following Figure 34, and pulse length may be obtainable from considering input activity level.

LIST OF REFERENCES

1. Shipley, M.N.,"The Application of Neural Networks to Acoustic Transient Classification", **Master's Thesis**, Naval Postgraduate School, Monterey, CA., Mar 92.

2. Urick, R.J., *Principles of Underwater Sound*, ed. 3, McGraw-Hill Book Co.,1983.

3. Hecht-Nielsen, Robert., *Neurocomputing*, Addison-Wesley Publishing Co., 1990.

4. Rumelhart, D. E., and McClelland, J. L., *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, v.1, MIT Press, 1986.

5. Neural Ware Inc., *Reference Guide for Professional II Plus*, Technical Publishing Group, Symatec Corp., Pittsburgh, PA. 1991.

6. Wann, M.,"Neural Network Investigation For ASW", Naval Air Warfare Center Publication, May 1992.

7. Barga, R.S., Friesel, M.A., Melton, R.B., "Classification of acoustic emission waveforms for nondestructive evaluation using neural networks", Proceedings of the International Society for Optical Engineering: Applications of Artificial Neural Networks, vol. 1294, April 1990.

8. Kinsler, L.E., Frey, A.R., Coppens A.B., Sanders, J.V., *Fundamentals of Acoustics*, ed. 3, John Wiley & Sons, 1982.

9. Friedlander, Benjamin., Transient Signal Detection Techniques., U.S. Navy Journal of Underwater Acoustics, Vol. 40 No. 4, October 1990

10. Devore, J.L., *Probability and Statistics for Engineering and the Sciences*, ed. 3, Brooks/Cole Publishing Company, 1991

INITIAL DISTRIBUTION LIST

		No. of Copies
1.	Professor D.J. Collins (Code AA/Co) Naval Postgraduate School Monterey, CA 93943-5000	1
2.	Professor O.B. Wilson (Code PH) Naval Postgraduate School Monterey, CA 93943-5000	3
3.	Professor K.E. Woehler (Code Ph/Co) Naval Postgraduate School Monterey, CA 93943-5000	1
4.	Naval Undersea Warfare Center New London Detachment New London, CT 06320	1
5.	Commander Space and Naval Warfare Center Dept. of Navy Washington, D.C. 20363-5100	1
6.	USS Puffer SSN 652 ATTN: LCDR T. S. Wichert FPO AP 96675-2338	1
7.	Library (Code 52) Naval Postgraduate School Monterey, CA 93943-5002	2
8.	Defense Technical Information Center Cameron Station Alexandria, VA 22304-6145	2

Thesis
W5725 Wichert
c.1 Feature based neural
network acoustic transient
signal classification.